SYNTHESE LIBRARY 341

Sten Lindström
Erik Palmgren
Krister Segerberg
Viggo Stoltenberg-Hansen
Editors

Logicism, Intuitionism, and Formalism

What Has Become of Them?

Springer

LOGICISM, INTUITIONISM, AND FORMALISM

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,

LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

VOLUME 341

LOGICISM, INTUITIONISM, AND FORMALISM

WHAT HAS BECOME OF THEM?

Edited by

Sten Lindström
Umeå University, Sweden

Erik Palmgren
Uppsala University, Sweden

Krister Segerberg
Uppsala University, Sweden

and

Viggo Stoltenberg-Hansen
Uppsala University, Sweden

Springer

Editors
Prof. Sten Lindström
Umeå University
Dept. Historical, Philosophical
and Religious Studies
SE-901 87 Umeå
Sweden
sten.lindstrom@philos.umu.se

Prof. Erik Palmgren
Uppsala University
Department of Mathematics
Box 480
751 06 Uppsala
Sweden
erik.palmgren@math.uu.se

Prof. Krister Segerberg
Uppsala University
Department of Philosophy
Box 627
751 26 Uppsala
Sweden
krister.segerberg@filosofi.uu.se

Prof. Viggo Stoltenberg-Hansen
Uppsala University
Department of Mathematics
Box 480
751 06 Uppsala
Sweden
viggo.stoltenberg-hansen@math.uu.se

ISBN: 978-1-4020-8925-1 e-ISBN: 978-1-4020-8926-8

DOI 10.1007/978-1-4020-8926-8

Library of Congress Control Number: 2008935522

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

The present anthology has its origin in two international conferences that were arranged at Uppsala University in August 2004: “Logicism, Intuitionism and Formalism: What has become of them?” followed by “Symposium on Constructive Mathematics”. The first conference concerned the three major programmes in the foundations of mathematics during the classical period from Frege’s *Begriffsschrift* in 1879 to the publication of Gödel’s two incompleteness theorems in 1931: The logicism of Frege, Russell and Whitehead, the intuitionism of Brouwer, and Hilbert’s formalist and proof-theoretic programme. The main purpose of the conference was to assess the relevance of these foundational programmes to contemporary philosophy of mathematics. The second conference was announced as a satellite event to the first, and was specifically concerned with constructive mathematics—an active branch of mathematics where mathematical statements—existence statements in particular—are interpreted in terms of what can be effectively constructed. Constructive mathematics may also be characterized as mathematics based on intuitionistic logic and, thus, be viewed as a direct descendant of Brouwer’s intuitionism. The two conferences were successful in bringing together a number of internationally renowned mathematicians and philosophers around common concerns. Once again it was confirmed that philosophers and mathematicians can work together and that real progress in the philosophy and foundations of mathematics is possible only if they do. Most of the papers in this collection originate from the two conferences, but a few additional papers of relevance to the issues discussed at the Uppsala conferences have been solicited especially for this volume.

Many people have helped us in making the two conferences and this volume possible. The person who has meant the most from a scientific point of view is Professor Per Martin-Löf whose vision and good judgement inspired and accompanied us through the preliminary stages, and who through his personal involvement during the conferences contributed to the positive result.

The Conference on the Philosophy of Mathematics was organized by the Department of Mathematics and the Department of Philosophy at Uppsala University in cooperation with the Swedish National Committee for Logic, Methodology and Philosophy of Science. The Symposium on Constructive Mathematics was organized by the Department of Mathematics. We are grateful to the two departments for financial support and, especially, to the Department of Mathematics for supplying the venue and for their generous organizational support. In this connection, the untiring organizational work of Zsuzsanna Kristófi was essential for the success of both conferences. We also wish to thank Ryszard Sliwinski for organizing things from the philosophy side. Moreover the help of PhD students Fredrik Dahlgren, Olov Wilander and Johan Granström with practical publicity issues was much appreciated. The latter also transformed Martin-Löf's manuscript into LaTeX.

We are also very grateful to the Swedish Research Council (VR) and the Royal Swedish Academy of Sciences (KVA) for financial support. Sten Lindström's editorial work on this volume has been made possible by a research grant ("The Ontology and Epistemology of Mathematics") from the Bank of Sweden Tercentenary Foundation (RJ) and from a fellowship during 2007 – 08 at the Swedish Collegium for the Social Sciences (SCAS). Erik Palmgren's editorial work was supported by a grant from the Swedish Research Council.

All papers in this volume, except the few that are reprinted, were assigned anonymous referees by the editors. We wish to thank these referees for their excellent work. We are also grateful to Springer's anonymous referee of the entire volume for insightful and helpful comments.

Finally, we wish to thank Professor Vincent Hendricks, Editor-in-Chief of Synthese Library, for encouraging this project; and Floor Oosting and Ingrid van Laarhoven at Springer, and Indumadhi Srinivasan at Integra Software Services for all their help in connection with the production of this book."

Umeå, Sweden
Uppsala, Sweden
October 2008

Sten Lindström
Erik Palmgren
Krister Segerberg
Viggo Stoltenberg-Hansen

Notes on Contributors

Peter Aczel is Professor of Mathematical Logic and Computing Science at Manchester University. His main research interests at present are in the foundations of mathematics and in constructive mathematics, particularly constructive set theory and constructive type theory.

Mark van Atten is researcher for the Centre National de Recherche Scientifique (CNRS) at the Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST) in Paris.

Hourya Benis Sinaceur has taught Logic and Philosophy of Science at the University Paris 1-Sorbonne. She is at present working at the CNRS/Institut d'Histoire et de Philosophie des Sciences et des Techniques in Paris. Recent publications include: *Corps et Modèles. Essai sur l'Histoire de l'Algèbre Réelle* (Paris, Vrin, second ed.: 1999); the edition of Alfred Tarski's 'Address at the Princeton University Bicentennial Conference on Problems of Mathematics' (The Bulletin of Symbolic Logic, March 2000); the translation into French of Paul Bernays' Abhandlungen zur Philosophie der Mathematik, Wissenschaftliche Buchgesellschaft, Darmstadt, 1976 (Paris, Vrin, 2003).

Josef Berger obtained his PhD on 'Applications of model theory to stochastic analysis' from the University of Munich in 2002. Since then his main interest is constructive mathematics. Currently he is a postdoc at the Japan Advanced Institute of Science and Technology, where he is working on constructive reverse mathematics.

Douglas Bridges is Professor of Pure Mathematics at the University of Canterbury, New Zealand. He has worked for thirty-five years in constructive analysis, topology, and foundations, with a side interest in mathematical economics, and has published over 140 papers and seven books. The latter include the monograph *Constructive Analysis* (with the late Errett Bishop) and the recent book *Technique of Constructive Analysis*, co-authored with Luminiţa Simona Vîţă. He holds D.Phil. and D.Sc.

degrees from the University of Oxford, is a Fellow of the Royal Society of New Zealand, and is a Corresponding Fellow of the Royal Society of his native city of Edinburgh.

John P. Burgess joined the philosophy department at Princeton University shortly after receiving his Ph.D. in logic from Berkeley, and has remained there ever since. He is the author of many papers in set theory and philosophical logic and philosophy of mathematics, and of the books *A Subject with No Object* (with Gideon Rosen) and *Fixing Frege* and the forthcoming *Mathematics, Models, and Modality: Selected Philosophical Papers* and *Philosophical Logic*.

Hajime Ishihara is an Associate Professor in the School of Information Science at Japan Advanced Institute of Science and Technology. He received his Ph.D. from Tokyo Institute of Technology in 1990. His research interests include constructive mathematics and its foundations, mathematical logic, and computability and complexity theory.

Juliette Kennedy received her Ph.D. in 1996 from the C.U.N.Y. Graduate Center (Department of Mathematics) with a thesis on models of arithmetic written under Attila Mate. After teaching at Stanford (1996–1997) and Bucknell (1997–1999) she moved to Finland, where she joined the Mathematics department of the University of Helsinki as, eventually, University Lecturer. She is now on leave from her job in Helsinki and visiting the Theoretical Philosophy group at the Utrecht Philosophy Department. Her interests, in no particular order, are technical, philosophical and historical, in the areas of, respectively, set-theoretic model theory, philosophy of mathematics and Gödel studies.

Sten Lindström is Professor of Philosophy at Umeå University and has been a Research Fellow at the Swedish Collegium for Advanced Study (SCAS). His main current research interests are in the philosophy of mathematics and philosophical logic. He has published papers on intensional logic, belief revision and philosophy of language, and co-edited the books *Logic, Action and Cognition: Essays in Philosophical Logic* (with Eva Ejerhed, Kluwer, 1997) and *Collected Papers of Stig Kanger with Essays on his Life and Work, I-II* (with Ghita Holmström-Hintikka and Rysiek Sliwinski, Kluwer, 2001).

Øystein Linnebo is a Lecturer in Philosophy at the University of Bristol, having held research positions at Oxford and the University of Oslo. He obtained a PhD in Philosophy from Harvard University in 2002 and an MA in Mathematics from the University of Oslo in 1995. Linnebo's main research interests are in the philosophies of logic and mathematics, metaphysics and the philosophy of language. His views are often inspired by those of his philosophical hero, Gottlob Frege.

Per Martin-Löf is Professor of Logic, Departments of Mathematics and Philosophy, University of Stockholm, Sweden.

Peter Pagin received his PhD in philosophy at Stockholm University in 1987, supervised by Dag Prawitz. He has since published papers on general meaning theory, compositionality, semantic holism, the semantics-pragmatics relation, reference and modality, assertion, synonymy, analyticity and indeterminacy, and philosophical aspects of intuitionism, among other things. He is currently Professor of Philosophy at Stockholm University.

Erik Palmgren is Professor of Mathematics at Uppsala University. His research interests are mainly mathematical logic and the foundations of mathematics. He is presently working on the foundational programme of replacing impredicative constructions by inductive constructions in mathematics, with special emphasis on point-free topology and topos theory.

Michael Rathjen is Professor of Mathematics at the University of Leeds. His main research area is mathematical logic, especially proof theory, type theory, and constructive set theory.

Peter Schuster is Privatdozent at the University of Munich. While his mathematical interests include constructive set theory, point-free topology, and formalisation in algebra, his related foundational focus is on the pretended necessity of higher-type ideal objects in mathematics.

Helmut Schwichtenberg is Professor of Mathematics at LMU Munich. His research areas are proof theory, lambda calculus, recursion theory and applications of logic to computer science.

Krister Segerberg is Emeritus Professor of Philosophy at Uppsala University and the University of Auckland. He is the author of papers in modal logic, the logic of action, belief revision and deontic logic, as well as the books *An Essay in Classical Modal Logic* (1971) and *Classical Propositional Operators: An Exercise in the Foundations of Logic* (1982).

Stewart Shapiro is currently the O'Donnell Professor of Philosophy at The Ohio State University, and a Professorial Fellow at the Arché Research Centre at the University of St. Andrews. His research interests include the philosophy of mathematics, logic, philosophy of logic, and philosophy of language. Major publications include *Foundations without foundationalism: a case for second-order logic*, *Philosophy of mathematics: structure and ontology*, and *Vagueness in context*. He has three children, and lives with his wife of 32 years in Columbus Ohio, spending about two months each year at St. Andrews in Scotland.

Wilfried Sieg is Professor of Philosophy and Mathematical Logic at Carnegie Mellon University. He works in proof theory, philosophy and history of modern mathematics, computation theory and, relatedly, in the foundations of cognitive science. Most relevant to his paper in the current volume is his editorial work concerning Kurt Gödel, David Hilbert, and Paul Bernays. As to Hilbert, he is editing (with William Ewald, Michael Hallett, and Ulrich Majer) Hilbert's unpublished notes of lectures from the 1890s to the early 1930s, in geometry, physics, as well as arithmetic and logic.

Sören Stenlund is Emeritus Professor of Philosophy at Uppsala University. He is the author of *Language and Philosophical Problems* (Routledge 1990), and has published several other books and articles on various themes in the philosophies of language, logic and mathematics. Problems concerning the nature and history of philosophy are other themes dealt with in Stenlund's publications, some of which are available only in Swedish.

Viggo Stoltenberg-Hansen is Emeritus Professor of Mathematical Logic at Uppsala University. His main interests include computability and constructivity in mathematics.

Neil Tennant is Humanities Distinguished Professor in Philosophy at The Ohio State University. He is an Alexander von Humboldt Fellow, a Fellow of the Academy of the Humanities of Australia, and an Overseas Fellow of Churchill College, Cambridge. He is the author of *Natural Logic* (Edinburgh, 1978), *Philosophy, Evolution and Human Nature* (with F. von Schilcher: RKP, 1984), *Anti-Realism and Logic* (OUP, 1987), *Autologic* (Edinburgh, 1992) and *The Taming of The True* (OUP, 1997). His current research interests are the philosophy and foundations of mathematics, logic and belief-revision.

Wim Veldman completed his dissertation entitled *Intuitionistic Hierarchy Theory* under the guidance of Johan J. de Iongh in 1981. Since then he has been teaching various subjects in the foundations of mathematics, in particular intuitionistic mathematics, at the Radboud University Nijmegen (formerly: Katholieke Universiteit Nijmegen). In his research, he has been trying to further develop intuitionistic mathematics as envisaged by L.E.J. Brouwer.

Luminiţa Simona Vîţă obtained her undergraduate education at the University of Bucharest. After gaining her Ph.D. degrees both from Canterbury and Bucharest, she held Postdoctoral Research Fellowships supported by the Royal Society of New Zealand. She has published many papers in constructive analysis, an undergraduate book on computability and co-authored with Douglas Bridges a graduate text on constructive mathematics. She is currently a research economist at the New Zealand Institute of Economic Research in Wellington.

Contents

Introduction: The Three Foundational Programmes

Sten Lindström and Erik Palmgren

1 Overview

The period in the foundations of mathematics that started in 1879 with the publication of Frege's *Begriffsschrift* [18] and ended in 1931 with Gödel's [24] *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*[1] can reasonably be called the classical period. It saw the development of three major foundational programmes: the logicism of Frege, Russell and Whitehead, the intuitionism of Brouwer, and Hilbert's formalist and proof-theoretic programme. In this period, there were also lively exchanges between the various schools culminating in the famous Hilbert-Brouwer controversy in the 1920s. The state of the foundational programmes at the end of the classical period is reported in the papers by Carnap, Heyting and von Neumann (Cf. Benacerraf and Putnam [1]) from the Conference on Epistemology of the Exact Sciences in Königsberg 1930. This was the very same symposium at which Gödel announced his First Incompleteness Theorem.[2]

A purpose of this volume is to review the programmes in the foundations of mathematics from the classical period and to assess their possible relevance for contemporary philosophy of mathematics. What can we say, in retrospect, about the various foundational programmes and the disputes that took place between them? To what extent do the classical programmes of logicism, intuitionism and formalism represent options that are still alive today?

A set of papers in constructive mathematics were specially solicited to the present anthology. This active branch of mathematics is a direct legacy of Brouwer's intuitionism. Today one often views it more abstractly as mathematics based on

S. Lindström (✉)
Department of Historical, Philosophical and Religious Studies, Umeå University, Umeå, Sweden
e-mail: sten.lindstrom@philos.umu.se

[1] *On formally undecidable propositions of Principia Mathematica and related systems,* reprinted in van Heijenoort [29].

[2] Many important original papers of the period are contained in van Heijenoort [29]. See Hesseling [31] for a recent historical account of the relation between the programmes at that time. See also Mancosu [41] where many of the major articles from the foundational debate between intuitionists and formalists in the 1920s appear in English translation.

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_1,

intuitionistic logic. It can then be regarded as a generalisation of classical mathematics in that it may be given, firstly, the standard set-theoretic interpretation, secondly, algorithmic meaning, and thirdly, nonstandard interpretations in terms of variable sets (sheaves over topological spaces).

The volume will be of interest primarily to researchers and graduate students of philosophy, logic, mathematics and theoretical computer science. The material will be accessible to specialists in these areas and to advanced graduate students in the respective fields.

2 Logicism and Neologicism

2.1 Frege's Logicism and Dedekind's Analysis of the Natural Numbers

Kant claimed that our knowledge of mathematics is *synthetic apriori* and based on a faculty of *intuition*. Frege accepted Kant's claim in the case of geometry, i.e., he thought that our knowledge of Euclidian geometry is based on pure intuition of space. But he could not accept Kant's explanation of our knowledge of statements about numbers.

> The basis of arithmetic lies deeper, it seems, than that of any of the empirical sciences, and even than that of geometry. The truths of arithmetic governs all that is numerable. This is the widest domain of all; for to it belongs not only the existent [das Wirkliche], not only the intuitable [das Anschauliche], but everything thinkable. Should not the laws of number, then, be connected very intimately with the laws of thought?[3]

Frege thought of numerical statements as being objectively true or false. Moreover, he interpreted these statements as literally being about abstract mathematical objects that do not exist in space or time. Now the question arose: *How can we have knowledge about numbers and their properties, if numbers are abstract objects*? Clearly we cannot interact causally with abstract entities. Neither is it plausible to explain our knowledge of them in terms of some kind of non-empirical intuition. For Frege it was evident that knowledge about numbers is possible only if it is conceptual and apriori, rather than based on experience or intuition. Thus his main question became:

> How, then are numbers given to us, if we cannot have any ideas or intuitions of them?[4]

In order to show that apriori knowledge of arithmetic is possible, Frege thought it necessary and sufficient to establish *the logicist thesis* that arithmetic is reducible to logic. More precisely, he wanted to show that:

---

[3] Frege [19], Section 14.

[4] "Wir soll uns denn eine Zahl gegeben sein, wenn wir keine Vorstellung oder Anschauung von ihr haben können?", Frege, *Grundlagen der Arithmetik* [19], Section 62.

(i) the concepts of arithmetic can be explicitly defined in terms of logical concepts;
(ii) the truths of arithmetic can be derived from logical axioms (and definitions) by purely logical rules of inference.

The underlying epistemic idea seems to be that a rational subject could gain apriori knowledge of logic and the definitions of arithmetical concepts in logical terms. The same subject could then, in principle, infer the truths of arithmetic from the logical axioms and the definitions, and thereby gain apriori knowledge of arithmetical truths. In other words, for every arithmetical truth it is according to Frege possible for an ideally rational subject to gain apriori knowledge of that truth. Of course, a critic might question the assumption that the basic principles of Frege's logic are knowable apriori. Frege's reason for this assumption was presumably that he thought of these principles as conceptual truths. However, as it turned out, the principles of the logical system that Frege devised in *Grundgesetze* [20] were actually inconsistent.

The following four claims are implicit in Frege's logicist programme: (a) Logic is (or can be presented as) an interpreted *formal system* (a *Begriffsschrift*); (b) It can be known apriori that the axioms of logic are true and that the logical rules of inference preserve truth; (c) the concepts of arithmetic are logical concepts; and (d) the truths of arithmetic are provable in logic. From (a) and (b) it follows that the theorems of logic are true. Since a contradiction cannot be true, it follows that logic is consistent. Moreover, it seems to follow from (b) that we can gain apriori knowledge of the theorems of logic by proving them. In virtue of (d) then, arithmetic must be consistent and its truths knowable apriori.

Roughly at the same time as Frege conceived of his logicist program, Dedekind [12] was also arguing for a kind of logicism:[5]

> In science nothing capable of proof ought to be accepted without proof. Though this demand seems so reasonable yet I cannot regard it as having been met even in ... that part of logic which deals with the theory of numbers. In speaking of arithmetic (algebra, analysis) as a part of logic I mean to imply that I consider the number concept entirely independent of the notions of intuition of space and time, that I consider it an immediate result from the laws of thought. (Dedekind [12], quoted from the Engl. trans. *Essays on the Theory of Numbers*, p. 31)

Working within informal set theory (which he considered to be a part of logic), Dedekind gave—for the first time—an abstract axiomatic characterisation of the system of natural numbers. To be precise, he characterised the structure of the system of natural numbers (up to isomorphism) by means of the notion of a *simple infinite system*: A simply infinite system is a set X (representing the natural numbers)

[5] Dedekind's version of logicism is discussed in great detail, both systematically and from a historical point of view, in Sieg and Schlimm [51]. See also Reck [47], who aptly refers to Dedekind's approach as "logical structuralism".

together with an element e (representing 0) and an operation S in X (the successor function), satisfying the conditions:[6]

(a) S is a one-to-one mapping from X onto $X - \{e\}$.
(b) for any set $Y \subseteq X$, if $e \in Y$ and $S(x) \in Y$, whenever $x \in Y$, then $Y = X$. That is, X is the smallest set containing e and being closed under the successor operation.

In modern terms, such a system is a (standard) model of *Peano's axioms* (or more accurately the *Dedekind-Peano's axioms*[7]) for the natural numbers:[8]

(P1) 0 is a natural number.
(P2) Every natural number n has a unique successor $S(n)$.
(P3) 0 is not the successor of any natural number.
(P4) Two different natural numbers do not have the same successor.
(P5) For every F, if the following two conditions hold: (a) $F(0)$, and (b) for every natural number n, if $F(n)$, then $F(S(n))$, then for every natural number n, $F(n)$. (*The Principle of Mathematical Induction*)

Dedekind proved that any two simply infinite systems are isomorphic. This means that second-order *Peano arithmetic*, with the standard semantics, is *categorical*, i.e., all its models are isomorphic,[9] and hence it is negation-complete, i.e., for any sentence φ in the language of second-order arithmetic, either φ or $\neg\varphi$ is a *logically consequence* of Peano's axioms (i.e., true in all models of the axioms).

The natural number system Dedekind thought of as being obtained by a process of abstraction: Starting from any simply infinite system one abstracts from those features that distinguishes it from any other simply ordered system. One thereby obtain the abstract system of natural numbers that Dedekind describes as a "free creation of the human mind". In "Letter to Keferstein" [13], Dedekind speaks of the natural number sequence as the "abstract type" of simply infinite systems:

---

[6] Dedekind actually thought of the natural numbers as starting with 1 instead of 0, as is customary nowadays.

[7] These axioms were presented independently by Dedekind [12] and Peano [46]. See also Dedekind's brilliant discussion in "Letter to Keferstein" [13] from 1890 of the basic ideas underlying the axioms.

[8] We are here considering the second-order language of Peano arithmetic with its *standard* model-theoretic semantics. In each (standard) model, the unary predicate variables range over the entire power set of the domain D of individuals. And for $n > 1$, the n-ary predicate variables range over the power set of D^n. The standard semantics for second-order logic stands in contrast to the "non-standard" semantics devised by Henkin [30], which in addition to standard models, also allows for *generalised* or *Henkin models*. A *generalised model* is a structure where the unary predicate variables range over some non-empty subset of the full power set of D, and similarly for the n-ary predicate variables. Validity and logical consequence are defined relative to all generalised models. Second-order logic with the Henkin semantics is recursively axiomatizable as well as compact and satisfies Löwenheim–Skolems theorem. Second-order logic with the standard semantics has none of these properties. See Shapiro [48] for a detailed development of second-order logic, as well as a defence of its use as a formal framework for mathematics.

[9] On the other hand, among the Henkin models for PA, there are also models containing non-standard numbers greater than all the "natural" numbers.

> After the essential nature of the simply infinite system, whose abstract type is the number sequence N, had been recognised in my analysis (articles 71 and 73), the question arose: does such a system exist at all in the realm of our ideas? Without a logical proof of existence it would always remain doubtful whether the notion of such a system might not contain internal contradictions. Hence the need for such proofs (articles 66 and 72 of my essay).

Dedekind defined a set X to be *infinite* if it can be mapped 1-1 and onto a proper subset of itself.[10] He, then showed that any infinite set includes a simply infinite system. In other words, the existence of an infinite set is necessary and sufficient for the existence of a simply infinite system. And the existence of a simply infinite system is necessary for the semantic consistency (i.e., satisfiability) of second-order Peano arithmetic. As we see from the quote above, Dedekind wanted to give a logical proof of the existence of a simply infinite system (or equivalently of a Dedekind-infinite set). In modern set theory one simply takes it as an axiom, the *Axiom of Infinity*, that there is an infinite set. But this procedure is questionable from a logicist point of view, since it is far from obvious that such an axiom is logically true.

Dedekind in fact thought that he could prove that there exists an infinite set.[11] In his "proof" he considers his "own realm of thoughts", i.e., "the totality S of all things, which can be object of my thought" and argues that it must be infinite. For any x in S, he defines the successor $s(x)$ as the thought that x can be an object of my thought. This thought, he maintains, can be an object of my thought, hence for any $x, s(x) \in S$. Moreover, there are elements in S that are not themselves thoughts "e.g., my own ego". Furthermore, Dedekind argues the mapping s is one-to-one. Thus s is a one-to-one mapping from S into a proper subset of S. Hence, S is Dedekind-infinite.

Dedekind's "proof" of the existence of an infinite set, lacks the stringency that one would ordinarily expect from a mathematical proof. As soon as one realises that every plurality cannot be assumed to form a set (e.g., there is presumably no set of all non-self-membered sets), then one sees that Dedekind has not proved that there actually *exists* a set S of all things that can be object of "my" thought. For all that Dedekind says, it is quite possible that there is no such set.[12] Hence, Dedekind's

---

[10] This is the notion of a set being *Dedekind-infinite*. The notion of Dedekind-infinite does not presuppose the notion of natural number. In Zermelo-Fraenkel set theory, **ZF**, it is provable that every Dedekind-infinite set is *infinite*, in the standard sense of not being finite, i.e., equinumerous to an initial segment of the natural numbers. The converse, i.e., that every infinite set is Dedekind-infinite is not provable in **ZF**, but it is provable in **ZFC** (Zermelo-Fraenkel set theory with the Axiom of Choice).

[11] Cf. Dedekind [12], Theorem 66.

[12] Indeed, the assumption that there is such a set is threatened by paradox. Suppose, namely, that the set S exists. Then, presumably, the subset R of S consisting of all the objects in S that are not members of themselves also exists. Thus, for any member x of S, x belongs to R just in case it does not belong to itself. But it seems that R itself can be an object of my thought. Hence, R is a member of S. Consequently, R belongs to R if and only if it does not. As a matter of fact, in "Letter to Dedekind" [9] Cantor gives the "totality of everything thinkable" as an example of an *absolutely infinite* or *inconsistent* multiplicity that cannot be "gathered together" into "one thing", i.e., that does not form a set.

attempted proof that there is an infinite totality consisting of all things that can be objects of "my" thought must be considered a failure.

To prove that there are infinitely many objects—or for that matter, to prove that an infinite set exists—is a major stumbling block for any logicist program. From a modern perspective it is hard to see, for example, how the existence of infinitely many natural numbers could ever be provable within logic. According to the dominant contemporary view, logic should be topic-neutral and free of ontological commitment.[13] As it seems, logicism demands a completely different conception of logic than the standard contemporary one.[14]

Before embarking on a logicist programme one ought to have some idea about how to answer questions of the following kind: (i) What are the criteria for distinguishing *logical concepts* from non-logical ones? (ii) What is it that makes an axiom or a rule *logical*? (iii) How can we recognise something as a logical axiom or rule? The mark of the logical concepts and the logical laws for Frege seems to have been their complete generality. This criterion, however, appears to be too vague to be workable. Even today the problem of characterising logical concepts and logical principles is wide open.

A sentence (of second-order arithmetic) is *true* (simpliciter) if and only if it is true in the intended model i.e., the structure $< N, 0, S >$ consisting of the natural numbers N, the number 0, and the successor operation S on natural numbers. But this holds if and only if the sentence is true in all the standard models of second-order Peano arithmetic (since these models are all isomorphic). In other words, a sentence in the language of (second-order) Peano arithmetic is true if and only if it is a (standard) model-theoretic consequence of the Dedekind-Peano axioms. So it might appear that Frege's logicist programme would be accomplished if one could define the basic concepts of Peano arithmetic within logic and derive the Dedekind-Peano axioms from logical axioms and definitions by means of logical rules of inference. This was in essence what Frege wanted to do. For this purpose he invented a formal system of higher-order logic (including a theory of extensions of concepts) and proved within it axioms equivalent to the Dedekind-Peano axioms.

Now, in hindsight, we know that Frege's logicist programme, in its original form, *could not* have succeeded. It is not just that the logic he actually used turned out to be inconsistent. In view of Gödel's first incompleteness theorem, there is no formal system that proves exactly those sentences of second-order arithmetic that are true. For any consistent formal logic the logicist may construct, there will be true arithmetical sentences that are not provable in it. The modern neo-Fregeans must, therefore, abandon Frege's impossible dream of showing that all arithmetic truths are formally provable in logic. Instead they must be satisfied with something weaker.

---

[13] Standard predicate logic is ontologically committed to the existence of at least one object since the domain of quantification is required to be non-empty. This relatively harmless ontological commitment can be eliminated at the price of some loss of elegance.

[14] See, for example, Goldfarb [23] concerning the differences between Frege's conception of logic and the contemporary one.

2.2 Frege's Logic and Russell's Paradox

The formal logic that Frege actually used in [20] to carry out his programme consisted of the following ingredients:

(i) A *language* of higher-order predicate logic, where the individual variables are assumed to range over the collection of absolutely all objects and the variables of higher types are taken to range over "unsaturated" entities, i.e., *functions* and *concepts* (i.e., functions from entities of some kind to truth-values).

(ii) *Axioms* and *rules of inference* for higher-order predicate logic.

(iii) *Principles of Comprehension* for functions and concepts. For instance, there is a comprehension schema:

$$\exists F \forall x (F(x) \leftrightarrow A(x))$$

to the effect that every formula $A(x)$ (in which the variable F does not occur free) defines a first-level concept, i.e., a concept taking objects as arguments. The formula $A(x)$ is here allowed to be *impredicative*, i.e., it may involve quantification over all (first-level) concepts, including the concept that the formula defines.

(iv) *Frege's Basic Law V* according to which every concept F is associated with an object $\{x : F(x)\}$ (called the *extension* of F) satisfying the requirement that any two concepts F and G have the same extension just in case they are true of the same objects. That is:

$$(\textit{Basic Law V})\ \{x; F(x)\} = \{x : G(x)\} \leftrightarrow \forall x (F(x) \leftrightarrow G(x)).$$

The object $\{x : F(x)\}$ may be understood as the class of all object that fall under the concept F.

At first appearance, it might seem that the statement that every concept has an extension is a conceptual truth. After some hesitation, Frege also came to regard it as such. However, Basic Law V leads to a contradiction within Frege's system. The relation of coextensionality between concepts is an equivalence relation, and hence partitions the concepts into mutually exclusive equivalence classes. Intuitively, these equivalence classes represent classes of objects. In view of Cantor's theorem, there are more equivalence classes of concepts than there are objects. But Basic Law V is tantamount to assuming that there is a one-to-one mapping from equivalence classes of concepts to objects. Hence, there cannot be more equivalence classes of concepts than there are objects. We have a contradiction. In Fine's terminology [17], Basic Law V is *inflationary*, since it demands that there be more abstract objects representing concepts than there are objects. Now, Russell showed that this paradoxical argument can in fact be represented within Frege's system (*Russell's paradox*).

This is how Russell's paradox arises in Frege's system. In terms of the class abstraction operator $\{x : F(x)\}$, which Frege took as a primitive, one can define the membership relation $\in$:

$$x \in y \leftrightarrow \exists F[F(x) \wedge y = \{x : F(x)\}],$$

i.e., x is a member of the class y if and only if there is a concept F such that x falls under F and y is the class of all objects falling under F.

Within Frege's higher order logic, one can infer the following *Naive Comprehension Principle* for classes from Frege's Basic Law V:

$$\forall F \exists y \forall x (x \in y \leftrightarrow F(x)).$$

We say that an object x is *Russellian* (in symbols, $R(x)$) iff x is not a member of x (i.e., it is not the case that x is the extension of a concept F under which x itself falls). The existence of the concept R is guaranteed by the strong comprehension principles that are built into Frege's logic. Notice also that the definition of the concept R is *impredicative*, since it involves quantification over all concepts.

Since every concept has an extension, there is a class r (the *Russell class*) of all the objects that fall under the concept R, i.e., $r = \{x : x \notin x\}$. We can now easily derive a contradiction in Frege's system:

$$r \in r \leftrightarrow r \notin r.$$

Hence, Frege's logical system in [20] is inconsistent and thereby useless as a foundation for arithmetic.

Although Frege's programme in its original form cannot be carried out, it may be instructive to analyse what went wrong and see what can be done to repair Frege's system (Cf. Burgess [8]). One can obtain consistent subsystems of Frege's logic by modifying one or more of the ingredients of Frege's system that led to Russell's paradox. Thus, one can:

(i) Replace classical logic with a weaker one. One radical alternative is to choose a paraconsistent logic that tolerates contradictions by preventing them from trivialising the system. In such a system, not every sentence follows from a contradiction.

(ii) Replace Frege's impredicative principles of comprehension for concepts by weaker predicative ones. Then, one cannot prove that the Russell concept R or other impredicatively defined concepts exist. The mathematics that one obtains in this way is however very weak, even if one adopts Frege's axiom V of class abstraction (Cf. Burgess [8], Chapter 2).

(iii) Abandon or weaken Frege's theory of extensions. Here there are various alternatives: (a) base mathematical theories on a Neo-Fregean theory of extensions, where certain concepts are not assumed to have corresponding extensions; or

(b) base mathematical theories on Fregean abstraction principles (e.g., *Hume's principle* (see below) as a basis for Peano arithmetic).

Fregean approaches are characterised by an effort to make only minimal changes to Frege's original system in order to obtain consistency. One can then investigate how much of mathematics can be captured in the resulting systems. Fregean approaches should be contrasted with approaches that use, for example, (ramified or simple) type theory, axiomatic set theory, or category theory, as foundational frameworks for mathematics.[15]

2.3 Neo-Fregean Logicism

The new development started with Wright's book *Frege's Conception of Numbers as Objects* [56], where a proof is outlined of what has become known as *Frege's theorem*, namely that the standard axioms of arithmetic are provable in second-order logic extended with Hume's principle (so-called *Frege arithmetic*).[16] Hume's principle says that two concepts F and G have the same cardinal number iff they are equinumerous, i.e., iff there is a one-to-one correspondence between the objects falling under F and the objects falling under G. In symbols:

$$\text{(HP)}\ \forall F \forall G[NxFx = NxGx \leftrightarrow F \approx G],$$

where $F \approx G$ means that there exists a one-to-one correspondence between the objects that fall under F and G respectively.

According to Hume's principle, the concept of (*cardinal*) *number* is obtained by (Fregean) *abstraction* from the concept of *equinumerosity* between concepts (or properties). The latter concept is definable in second-order logic and is therefore, according to Frege and the neo-Fregeans, a logical concept. In second-order logic, Hume's principle becomes:

$$\begin{aligned}\forall F \forall G[NxFx = NxGx \leftrightarrow \exists R \forall x((Fx \to \exists! y(Gy \wedge R(x,y)))\\ \wedge(Gx \to \exists! y(Fy \wedge R(y,x))))].\end{aligned}$$

---

[15] The logicism of Whithead and Russell's *Principia Mathematica* [55] differs radically from Frege's in abandoning the idea that classes and numbers are objects. Propositions which, on the face of it, speak of classes or numbers are analyzed as really being about higher-order entities, namely, in Russell's terminology, *propositional functions*. See Linsky [40] for a thorough analysis of the logical framework of *Principia Mathematica*, i.e., Russell's Ramified Theory of Types, with its three characteristic axioms of Choice, Infinity, and Reducibility.

[16] That Frege's theorem is implicit in Frege's own work was first pointed out in 1965 by Charles Parsons [45, 14]. Subsequent reconstructions of Frege's logic by Boolos and Heck (Boolos [3, 4], Heck [27, 14]) confirm that Frege's theorem was indeed proved by Frege himself in [20], although he did not state it explicitly.

Wright and Hale avoid using Frege's inconsistent theory of classes or axiomatic set theory as a foundation for mathematics. Instead they use abstraction principles like Hume's principle to introduce various domains of mathematical entities. Hume's principle is taken as an implicit definition of the cardinal number operator: the number of x such that: $F(x)$, or in symbols: $NxF(x)$. Intuitively, $NxF(x)$ is the cardinal number associated with the concept F, i.e., the number of objects falling under the concept F. From this operator, one can define the natural numbers: $0 = Nx(x \neq x)$, i.e., the number of objects that are not self-identical. $1 = Nx(x = 0)$, $2 = Nx(x = 0 \vee x = 1)$, etc. Frege shows how one can define the concepts of successor and natural number (finite cardinal) in Frege arithmetic. Using Frege's definitions and Hume's principle, one can prove the Dedekind-Peano axioms for the natural numbers including the (second-order) *Principle of Mathematical Induction*.

The idea of using Hume's principle as a logical basis for arithmetic goes back to Frege's *Grundlagen der Arithmetik* [19], but was immediately rejected by Frege with the motivation that Hume's principle is not strong enough to uniquely identify the cardinal numbers. Hume's principle answers to the question under what conditions two cardinal numbers are identical. But it does not tell us what objects are cardinal numbers. For instance, it does not say whether the number 2 is identical to the class of all classes that are equinumerous to {0, 1} as Frege thought, or whether it is identical to the class {0, 1} itself, as von Neumann suggested. It cannot even answer the question whether Julius Caesar is a number (The *Julius Caesar Objection*).[17] For this reason, Frege preferred defining the natural numbers explicitly in terms of classes rather than basing arithmetic on Hume's principle. Another reason for Frege not to base arithmetic on Hume's principle, was presumably that he did not think that it could be taken as a fundamental principle of logic, since its subject matter was too specific for that purpose.

Wright and Hale ([56], [26]) advocate Hume's principle as a basis for arithmetic and argue that the principle is analytically true. The truth of the principle in turn implies the existence of infinitely many natural numbers. This means that they abandon Frege's view that arithmetic is in a strict sense reducible to logic, while retaining the idea that our knowledge of arithmetic is apriori and based on analytic truths. The neo-Fregean programme also aims at defining other domains of mathematical entities by means of abstraction principles, for instance, the domain of real numbers, and the domain of sets. A difficult problem is to differentiate between "bad" abstraction principles (cf. Frege's Basic Law V) that are inconsistent and "good" ones that can be the basis of mathematical concept formation. If there is no principled way of making such a distinction, it might seem that all abstraction principles are more or less suspect ("the Bad Company Objection").[18]

---

17 Cf. Heck [28].

18 Boolos [3] has given a relative consistency proof to the effect that Frege arithmetic is consistent provided that second-order Peano arithmetic is. See Fine [17], for a comprehensive study of the philosophical and logical aspects of Fregean abstraction and "the Bad Company Objection".

2.4 Fregean Abstraction Principles

Basic Law V and HP are examples of (Fregean) abstraction principles, i.e., principles of the form:

$$\$F = \$G \leftrightarrow F \sim G,$$

where $\sim$ is an equivalence relation between concepts and \$F and \$G are objects representing "equivalence classes" of concepts with respect to the relation $\sim$. A Fregean abstraction principle may be viewed as postulating the existence of a mapping \$ from equivalence classes of concepts to objects. In particular, Frege's Basic Law V postulates the existence of such a mapping from equivalence classes of concepts to objects that is one-to-one. This means that there must exist at least as many objects as there are concepts. On the other hand, the strong axioms of comprehension for concepts imply that there are more concepts than there are objects. Thus, we get a contradiction. HP is also a strong assumption, but it does not, as far as one knows, lead to inconsistency. In the context of second-order logic with unlimited comprehension principles for concepts, HP implies that there are infinitely many objects, i.e., any model of Frege arithmetic has to be at least denumerably infinite. Thus, Fregean abstraction principles can be very powerful as HP, or even inconsistent as Frege's Basic Law V.

Wright and Hale argue that Frege's theorem is of great philosophical importance. That is, they think that HP can be viewed as an implicit definition of the concept of a cardinal number, and therefore as an analytic truth. Given the truth of HP, one can prove that infinitely many cardinal numbers exist: Consider the empty concept $[x{:}x \neq x]$ of being an object which is not identical with itself. This concept exists by concept comprehension. Let E be this concept. Then, the formula $E \approx E$ is logically true. Hence, by HP the following is true: $NxE(x) = NxE(x)$. Now, according to the Fregean semantics, adopted by Hale and Wright, this sentence can be true only if the singular term $NxE(x)$ refers to an object. Hence there must exist some object that is the number of things that are not self-identical. But this object is by definition the cardinal number 0. Once 0 has been proved to exist, one can prove the existence of $1 = Nx(x = 0)$, in a completely analogous way, ... Thus, Wright and Hale claim that it is analytically true and apriori that there are infinitely many cardinal numbers.

The "Scottish neo-logicists" Hale and Wright do not argue that Hume's principle is a logical truth. Instead they claim that it (or, some modified version of it) is an analytic truth concerning the concept of a cardinal number. Hence, they give up Frege's idea of a strict reduction of arithmetic to logic, while keeping the Fregean doctrine that arithmetic has a foundation that is analytical and apriori. It is part of their neo-logicist programme to try to show that also other areas of mathematics can be logically based on analytically true abstraction principles. The programme is based on the conviction that substantial portions of mathematics can in this way be shown to be analytically true and apriori.

2.5 Contributions to this Volume

This book contains four papers, by John Burgess, Øystein Linnebo, Neil Tennant, and Stewart Shapiro, discussing various aspects of logicism and neo-logicism.

Burgess investigates the attenuated form of logicism that was introduced by Richard Jeffrey under the name *logicism lite*. According to this view, mathematics, although not reducible to logic, is a theoretical superstructure built upon logic and testable against logical data that are given in the form of logically valid schemata. Thus, "logicism lite" can be seen as a kind of instrumentalism: mathematical theories are seen as instruments for efficiently drawing logically valid conclusions. On such a view, mathematical objects like numbers and sets are "useful fictions" and mathematical theories are neither true nor false. Although finding "logicism lite" technically interesting, Burgess rejects it as a general philosophy of mathematics, arguing that the type of instrumentalism and fictionalism that it represents is contrary to the realist attitude of the working mathematician.

Linnebo gives an explanation of our reference to the natural numbers along broadly Fregean lines, starting out from Frege's famous *Context Principle*, according to which a word has meaning only in the context of a proposition. On Linnebo's approach, the natural numbers are presented via numerals and questions about natural numbers are reducible to questions about numerals. Hence, the metaphysical status of natural numbers, Linnebo argues, is "thinner" than that of numerals.

Tennant's paper is concerned with completing the constructive logicist programme, which he started in [52], of deriving the Dedekind-Peano axioms within a theory of natural numbers that also accounts for their role in counting finite collections of objects. According to this approach, the primitive concepts of arithmetic are introduced via Gentzen-style introduction and elimination rules within a system of natural deduction for intuitionistic logic. In the present paper, the goal is to show how the axioms of addition and multiplication can be introduced in a conceptually satisfying way within such a constructive logicist approach. For this purpose, Tennant develops a natural deduction system for the logic of orderly pairing. Orderly pairing is here treated as a logical primitive with its own introduction and elimination rules. This notion is then used to formulate introduction and elimination rules for addition and multiplication.

Shapiro, finally, considers various motivations that have been given for logicist programmes in the foundations of mathematics: the rationale of such a programme may be (i) mathematical, i.e., to prove anything that is capable of being proved; (ii) to provide mathematics with a foundation that is *epistemologically secure*, (iii) to determine the *epistemological source* of our mathematical knowledge; or (iv) to find a foundation for mathematics that is *not in need of further justification*. Shapiro then uses these aims as "yardsticks" to take the measure of Wright's and Hale's programme of Fregean neo-logicism.

3 Intuitionism and Constructive Mathematics

We give a brief sketch of the emergence of intuitionism and constructivism in mathematics. For a comprehensive account of the history, see Hesseling [31], and for the mathematical aspects we refer to Troelstra and van Dalen [54] and Dummett [16]. Kronecker's criticism in the 1870s of Cantor's transfinite set theory is often considered as the starting point of the development. His finitistic standpoint regarding mathematical objects is condensed in his famous dictum "Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk".[19] This view was of course at odds with set theory which had stipulated the existence of an abundance of abstract infinite objects. The success of set-theoretic methods in the theory of functions soon became undisputed and they were vigourously employed by Weierstrass and others. In 1904 Zermelo surprised the mathematical world by proving that the set of real numbers could be well-ordered (cf. van Heijenoort [29]). He had used his version of the axiom of choice, a principle that was already implicit in Cantor's set theory. Among the critics of his method were the so-called French semi-intuitionists Baire, Borel and Lebesgue. Borel and Baire contended that only countable choices where admissible in mathematics. Such choices could be performed one by one using induction on natural numbers. Lebesgue went further in the criticism and claimed that choices were only admitted if they followed some law. Borel emphasised mathematics as a human activity rather than a formal one, subjective but communicable. Poincaré objected to another aspect of set-theoretic methods, namely the use of impredicative definitions. He also argued against Russell and the logicist reduction of mathematics to logic.

Brouwer started in his thesis from 1907 *Over de Grondslagen der Wiskunde* (*On the Foundations of Mathematics*), to analyse the situation in the foundations of mathematics.[20] He came to the conclusion that the use of logic as the basis of mathematics is unreliable. Instead it should be founded on (wordless) mental constructions and the intuition of time. Many of the views of the semi-intuitionists were shared by Brouwer. His analysis further concluded that it was not abstract mathematical objects that were the problem, but the unheeded application of the principle of excluded middle, in particularly when dealing with infinite objects (Brouwer [6]). He did not use formal logical language in his own writing. Intuitionistic logic was only formalised later by Heyting [32] in 1930; and partially already by Kolmogorov [37] in 1925. Also Brouwer's notion of mental construction was analysed and made precise in what is now called the *Brouwer-Heyting-Kolmogorov (BHK) interpretation*.

Brouwer's next important idea for intuitionism was his notion of a *choice sequence* (around 1917), which was obtained by a reflection on the intuition of time.

---

[19] *God made the integers, everything else is Man's work.*

[20] English translations of several of Brouwer's paper's can be found in van Heijenoort [29] and Mancosu [41]. See also Brouwer's *Collected Works* [7] edited by Heyting.

One may argue that a decision based on a given infinite proceeding sequence of data, integers or the like, must be taken after a finite length of time, unless one knows that the sequence obeys a certain law which makes it predictable. This has the consequence that any function of the totality, or "spread", of sequences must depend continuously on the data. A remarkable result is then that all functions from real numbers to real numbers are continuous. This contradicts classical mathematics, where a step function provides an immediate counter example. Brouwer would object that such a function is not defined at the actual step point, as this would require a survey of an entire infinite sequence of rationals that defines the point on the real line.

Other implications of his continuity principles are the *Fan Theorem* (FT) and the *Bar Theorem* (BT). The former is a reformulation of König's lemma and the second is a transfinite induction principle. Both are consistent with classical mathematics.

A computational and formalised model of the BHK-interpretation was provided by Kleene [36] with his so-called recursive realisability interpretation. This model refuted both FT and BT and showed that they had no immediate computational content in the sense of Church and Turing. On the other hand there are indirect constructive interpretations of choice sequences.

Kreisel and Troelstra (cf. Troelstra [53]) proved a conservation result that eliminated the need to use choice sequences in many circumstances. Models of choice sequences using topos theory were constructed by van der Hoeven and Moerdijk [34]. The use of the principles FT and BT in topology, e.g. as in the Heine-Borel theorem, can often by eliminated by the use of point-free spaces or locales. A crucial ingredient in the construction of such spaces is the inductive generation of covers, indeed also present in Kreisel's first sketch of the elimination theorem in 1968; see also (Martin-Löf [42], pp. 77–78).

Further precise mathematical versions of the BHK-interpretation were obtained by introducing various constructive type theories. Curry found a similarity between axioms for propositional logic and types for certain combinators in his combinatory logic (Curry and Feys [11]), Howard in 1969 (Howard [35]), building on this observation, constructed a type theory for predicate logic. A type theory suitable for a full development of constructive mathematics was devised by Martin-Löf ([43, 44]). The correspondence between types and formulas (or propositions) goes under the name of the *Curry-Howard isomorphism.*

A different and very influential interpretation of intuitionistic arithmetic, the so-called *Dialectica interpretation*, is due to Gödel and was published in the paper *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes*[21] from 1958 (Cf. Gödel [25]). This interpretation came to have a great influence on much later development of the revised Hilbert programme. The Dialectica interpretation was also based on a type theory. Type theory provided a new and more fundamental justification of the intuitionistic logical principles, and provides in a sense an

[21] On an extension of finitary mathematics which has not yet been used.

embodiment of Brouwer's mental constructions. In any case, the philosophy behind type theory clearly puts mathematical constructions before formal logic.[22]

Topos theory grew out of the work of Grothendieck on algebraic geometry in the late 1950s. The connection between this theory and model theory of intuitionistic logical systems became gradually clearer with the introduction of the elementary topos by Lawvere and Tierney 1969–1970 and Joyal's generalisation of Kripke- and Beth-semantics. Category theory, the basis of topos theory, reduces logic—just as type theory does—to very simple and basic mathematical constructions. This is not accidental as there are strong connections to the Curry-Howard isomorphism (Cf. Lambek and Scott [38]). Indeed, Lawvere suggested in [39] that category theory could be made the foundation of mathematics on par with set theory.

Prior to the introduction of abstract set-theoretic notions and the modern notion of function, there was no pressing need to consider a separate realm of constructive mathematics. Deliberate attempts to distinguish such a realm came only after these notions became widely used. Constructive mathematics in a wide sense includes finitism and French semi-intuitionism, where the logic is still classical, and Brouwer intutionism and Markov school constructive mathematics (from 1950s), in which the logic is intuitionistic. See [54] and [5] for surveys. An important legacy of intuitionism is so called *Bishop-style constructive mathematics*, which has developed into a largely independent mathematical field. The pioneer, Errett Bishop, shared much of Brouwer's anti-formalist and anti-logicist stance, in that mathematical constructions are considered to be prior to logic. Constructive mathematics, in a narrow sense, tends nowadays to be identified with the development initiated in Bishop's book *Foundations of Constructive Analysis* [2] from 1967. It has its roots firmly in Brouwer's intuitionism, and is indeed building on much of the results of his school. However it makes certain generalisations and more modest assumptions on the mathematical ontology. As a result it is intelligible from a classical set-theoretic viewpoint, as well from the viewpoint of computability, via recursive realisability or through generalised inductive definitions as in Martin-Löf type theory, or Aczel-Myhill constructive set theory.

3.1 Contributions to this Volume

We now turn to the individual contributed articles on intuitionism of the present volume. Martin-Löf analyses Zermelo's axiom of choice through an interpretation in a classical extension of a basic constructive type theory and identifies the cru-

---

[22] Modern constructive type theories such as the *Calculus of Constructions* [10] may also be considered as examples realising the logicist programme. For instance natural numbers may be defined by writing down a single type $N = (\Pi X)(X \to (X \to X) \to X)$, where ΠX denotes quantification over all types. All natural numbers and their properties can be constructed or derived, by applying abstraction and application to that type.

cial point where its justification is non-constructive. However, in the basic theory, the countable axiom of choice can actually be proved due to the meaning of the existential quantifier.

The mentioned axioms FT and BT make it possible to work freely with pointwise continuity, and derive, for instance, the Heine-Borel Theorem about uniform continuity on compact intervals. This volume contains two papers studying the role of FT. Berger and Schuster show in their paper that FT, in a restricted form, is equivalent to Dini's theorem. Veldman shows in his paper that FT is equivalent to a two-dimensional approximate version of Brouwer's fixed point theorem. These results are instances of so called reverse mathematics: deriving axioms from specific important mathematical theorems over a weak and constructive base theory. This makes it possible to judge exactly which axioms are needed to obtain a certain theorem.

The lack of numerical (or computational) meaning of the FT axiom led Errett Bishop, along with others, to consider uniformly continuous functions as the basic continuous maps. An obvious obstacle with this choice is that there must be a way of expressing uniformity, as indeed there is in metric or uniform spaces. Going beyond such spaces, various ideas of furnishing spaces with more information has been developed. In apartness spaces, a fundamental idea is that the assertion that points are apart is more informative than an assertion about equality: equality may in such spaces be defined from a basic apartness relation, whereas the converse is usually not possible constructively. Bridges and Vîţă, in their contribution to this anthology, give an overview of research in this area. Another approach is point-free topology, where the covering relation between neighbourhoods is the basic information. This allows for good notions of compactness and continuity, in fact, eliminating the need for axioms FT or BT. Two contributions on locale theory are included here. Aczel gives a constructive version of Lusin's separation theorem that avoids use of the BT axiom, employed in early intuitionistic versions of the theorem, with the help of point-free topologies. In the paper of Palmgren an application of point-free methods to the foundations of homotopy theory is given.

Various sharpenings of the constructive position in mathematics have been suggested over the years. One is to take into account not only computability, but also computational complexity. Ishihara's paper shows how a theory of real numbers can be developed on the assumption that their fundamental sequences of rational approximations are computable in a certain complexity class. Related to this is the approach in Schwichtenberg's paper, which refines notions of constructive analysis to make explicit witnesses to existence theorems. In addition he shows that a computer program for computing roots of functions may be extracted from a proof of the intermediate value theorem.

Dummett [15] considered the philosophical justifications for classical and intuitionistic logic from the meaning-as-use perspective. His conclusion is that the *principle of bivalence*, i.e. that every proposition is either true or false, has to be rejected as a logically valid principle, thus undermining the realist justification of classical logic. The debate of his anti-realist argument has been going on since then, with contributions by Prawitz among others. Dummett has argued that the rejection of bivalence—for meaning-theoretic reasons—has metaphysical consequences. According to this view, there are meaning-theoretic reasons for giving up a realist metaphysics.

Pagin has criticised Dummett's argument against metaphysical realism and argued that it contains a gap. The fact that bivalence has to be given up as a principle of logic—for meaning-theoretic reasons—does not imply that the principle cannot still be true. According to this line of reasoning, a realist metaphysics is not refuted by Dummett's meaning-theoretic arguments against the validity of bivalence. What is needed according to Pagin to close the gap between meaning theory and metaphysics is the principle:

(P) If A is true, then A is provable.

If (P) is provable, metaphysical realism (the *truth* of bivalence) will entail its provability. So, Dummett's meaning-theoretic argument against bivalence as a logical principle, would have metaphysical consequences after all. Pagin, however, has argued that the principle (P) is not intuitionistically acceptable and cannot be proved. Prawitz disagrees and has actually tried to prove a formal counterpart of (P). In his contribution to this volume, Pagin criticises Prawitz' argumentation. If there is no reason for holding on to (P), Pagin argues, there is no reason coming from meaning theory to doubt the truth of bivalence (i.e., metaphysical realism) either.

4 Formalism

A milestone in mathematics is Hilbert's *Grundlagen der Geometrie* [33] from 1899. Its importance for the conceptual development of modern mathematics is difficult to overstate. Here Hilbert gave, for the first time, a fully precise axiomatization of Euclidean geometry. The entities like point, line and plane are defined only implicitly by their mutual relations. Generalising this method of implicit definitions it became possible to work also with complicated mathematical systems characterised axiomatically up to structural equivalence or isomorphisms. Hilbert's structuralist approach, of course, goes back to Dedekind's characterisation in [12] of the natural number system in terms of simply infinite systems. It was also foreshadowed by Felix Klein's classification of geometries using group invariants (the Erlangen programme).

At the time of writing *Grundlagen der Geometrie*, Hilbert subscribed to the view that mathematical truth and existence simply means consistency. In a famous letter to Frege of December 29 1899, he wrote: "As long as I have been thinking, writing and lecturing on these things, I have been saying ... if the arbitrary given axioms do not contradict each other with all their consequences, then they are true and the things defined by them exist. This is for me the criterion of truth and existence." This structuralist approach of Hilbert made it possible for him to be indifferent to the ontological questions about the nature of mathematical objects that were of such great concern to Frege.[23]

The formalistic view of mathematics was shaken by the inconsistencies discovered in naive set theories, e.g. the Russell paradox and the Burali-Forti paradox. There were various cures proposed for the childhood illness of modern mathematics. One was to forbid impredicative definitions as prescribed by Poincaré, Borel and Russell. Another was Brouwer's, to restrict the use of classical logic. There were of course deep contentual considerations behind these suggestions.

4.1 Hilbert's Proof-Theoretic Programme

Throughout his career, Hilbert had a deep interest in foundational questions, although his views went through many changes. Around the turn of the century, Hilbert's foundational point of view was close to Dedekind's brand of logicism. Like Dedekind he thought that he could prove the consistency of fundamental mathematical theories like analysis and Euclidian geometry by constructing logical (or set-theoretic) models. This belief was shattered, however, by the discovery of the set-theoretic paradoxes. After that, Hilbert turned to the idea of providing fundamental mathematical theories, like arithmetic, analysis and set theory, with direct metamathematical proofs of consistency. In this way Hilbert thought that he could defend abstract infinitistic mathematics from constructivist critics like Kronecker, Brouwer and Weyl. In particular, he wished to respond to Brouwer's criticism of the unheeded use of the law of excluded middle in arguments about infinite objects. However, *Hilbert's proof-theoretic programme* took a long time to reach its mature form in the beginning of the 1920s. By taking a constructivist, or finitist, position, Hilbert attempted to provide a justification for abstract mathematics going way beyond the constructive basis. In spite of the heated "Grundlagenstreit" between Brouwer and Hilbert, and their followers, in the 1920s, Hilbert's program is best viewed as an attempt at mediating between classical and constructive mathematics.[24]

According to the formalist view, a consistent system, formalising a sufficiently rich ontology, was all that was required to carry out abstract mathematics, with all its ideal objects. Hilbert's idea was to prove the consistency of such a formal system in a finitistic system that Brouwer, or other constructivists, could not object to. This would save classical mathematics from constructivist criticism. For this purpose Hilbert devised his proof theory, which studies proofs of a formal system, and aim to show that no absurdity could ever be derived according to the rules of the system.

---

[23] See Frege [21] for the correspondence between Frege and Hilbert. See also Shapiro's paper "Categories, structures, and the Frege-Hilbert controversy: the status of metamathematics" in this volume for a philosophical analysis of Frege's and Hilbert's respective views on the role of mathematical axioms and the relationship between truth, consistency and mathematical existence.

[24] See Sieg's contribution to this volume [50] as well as Sieg [49] for a detailed analysis the development of Hilbert's foundational views. In this connection Zach [57] is also useful.

Proofs are used in mathematics to establish theorems. In *proof theory*, however, proofs, or rather formal derivations, are the object of mathematical study. Proof theory started with Hilbert's proposal to investigate mathematical proofs that involve abstract infinitistic reasoning by finitistic means. In particular, Hilbert proposed his *finitist consistency programme*: consider a formal system T in which all of classical mathematics can be formalised and prove by finitistic means the consistency of T. In this way, Hilbert wanted to prove the consistency of classical mathematics in a particularly elementary part: "finitistic mathematics". When Hilbert formulated his programme, he had two significant facts available:

(i) Classical mathematics can be represented in formal systems of set theory or type theory.
(ii) These formal systems can be described in a finitistic manner.

Although "finitistic mathematics" was not given a precise characterisation, it was clear that it was concerned with concrete spatio-temporal objects and that it was to employ only elementary ("finitistic") combinatorial methods. These methods were supposed to be epistemically secure and intuitively evident in a Kantian sense. It would then be possible, Hilbert thought, to find a common Kantian foundation for classical mathematics and intuitionistic mathematics.

This programme was shattered by Gödel's second incompleteness theorem which showed that consistency proofs of the kind Hilbert envisaged are not possible. A revised version of the Hilbert programme was started by Gentzen, in which the finitistic principles were extended. A novel principle was transfinite induction on a well-order defined by a finitistic notation system for ordinals. Such a system, indicated below, is supposed to be intuitively well-founded. Thus making the induction principle valid.

$$0, 1, 2, \ldots\ \omega, \omega + 1, \ldots\ \omega 2, \ldots\ \omega n, \ldots\ \omega^2, \ldots\ \omega^{\omega}, \ldots\ \omega^{\omega^{\omega}}, \ldots\ \varepsilon_0$$

Gentzen [22] proved the consistency of Peano arithmetic within a standard finitistic system extended with induction up to every ordinal number less than ε_0. This can of course only be a relative consistency proof. Similar consistency proofs have been carried out for much stronger theories. This branch of proof theory is still being developed, though a consistency proof for second order arithmetic seems unreachable at present. The search for ever stronger ordinal notation systems uses motivation and analogies from set theory and large cardinal axioms.

Another optimistic belief held by Hilbert was that every mathematical question was in principle decidable. This view seemed to have been brought down by Gödel's first incompleteness theorem. A possible way around this is to introduce a progression of ever more powerful systems, that would eventually decide a certain problem. The progression should naturally not be arbitrary, but driven by the intuition of the human mind. Gödel proposed to decide Cantor's Continuum Problem in this way, by postulating the existence of larger and larger cardinal numbers.

4.2 Contributions to this Volume

Category theory may be viewed as a modern implementation of structuralist ideas and is ubiquitous in mathematics of today. It takes these ideas even further into abstraction than, say, the Bourbaki school. This makes it possible to characterise classes of mathematical structures in terms of the mappings, or morphisms, between these structures leaving aside their internal workings. The structures need not have an underlying set as in the Bourbaki approach. Shapiro puts a current debate, on whether category theory itself needs a foundation and whether it is an appropriate framework for structuralism, into perspective by relating it to the controversy between Frege and Hilbert regarding the nature of axioms in mathematics and the relationship between the consistency of an axiom system and the existence of the mathematical entities it refers to.

Model theory, for instance first-order model theory, is another way of implementing structuralist ideas. Here the interplay of a restricted logical language and the models plays a critical role. Different relations between models may be considered that depend on this language being restricted. One can think here of the relation "elementary embedded in" and its role with regard to nonstandard models of arithmetic and real closed fields. Benis Sinaceur discusses Tarski's version of formalism in her paper.

Stenlund stresses the Kantian background to Hilbert's epistemological aims. He argues that the philosophical problems raised by Hilbert–in particular the problems surrounding the concept of infinity–are of great relevance today and should be taken seriously by philosophers of mathematics. The type of conceptual investigation required to solve these problems should not be conflated with technical work within formal logic and cannot be pursued by technical means.

Sieg gives a historical and philosophical exposition of Hilbert's programme. In this connection he argues against the popular view that Hilbert's programme was devised mainly as a response to Brouwer's criticism of the unheeded use of the law of excluded middle in arguments about infinite objects. Instead he stresses its historical roots in the debate between Dedekind and Kronecker and in the attempts, going back to Dedekind, of securing the consistency of modern abstract infinitistic mathematics. In the latter part of the paper, he considers modern extensions of Hilbert's foundational programme that would fit Bernays's and Hilbert's broader views on the foundations of mathematics.

Rathjen presents, in his contribution, a revision of Hilbert's programme from the viewpoint of type theory. The basic idea for this revision is contained in Gödel's Dialectica paper [25], where intuitionistic arithmetic is justified via the introduction of computable higher-type structures, and not via the assumption of long well-orders as in Gentzen's work.

The paper by van Atten and Kennedy is concerned with the incompleteness of ZFC (Zermelo-Fraenkel set theory with the Axiom of Choice) and related systems for set theory, and the way Gödel's thinking about incompleteness developed in his published and unpublished work. For Gödel the main philosophical question was:

> Are there *absolutely undecidable* questions in set theory, i.e., questions that are not only undecided in existing formal systems like ZFC, but cannot ever be decided by human reason?

According to Gödel's *set-theoretic platonism*, set theory describes a well-determined reality in which every set-theoretic statement is determinately either true or false. During much of his career Gödel seems to have shared Hilbert's "rational optimism" that it is in principle possible for the human mind to solve any mathematical problem; in particular, to answer the question whether the *Continuum Hypothesis* (CH) is true. In view of Cohen's result that CH is not provable in ZFC, a proof of CH would require new set-theoretic axioms. van Atten and Kennedy analyse the Gödelian project of deciding open questions in set theory by extending ZFC with new axioms, for instance, axioms postulating the existence of large cardinals. At the end of the paper, they consider the present situation in set theory in the light of Gödel's ideas.

Acknowledgments We are grateful to Wilfried Sieg for his very helpful comments on an earlier version of this Introduction.

References

1. Benacerraf, P., Putnam, H., *Philosophy of Mathematics: Selected Readings* (2nd ed.). Cambridge: Cambridge University Press, 1983.
2. Bishop, E., *Foundations of Constructive Analysis*. New York: Mac-Graw Hill, 1967.
3. Boolos, G., 'The Consistency of Frege's Foundations of Arithmetic', in J.J. Thompson (ed.), *Being and Saying: Essays for Richard Cartwright*. Cambridge: MIT Press, 1987. Reprinted in Boolos G., *Logic, Logic, and Logic*. Cambridge: Harvard University Press, 1998.
4. Boolos, G., *Logic, Logic, and Logic*. Cambridge: Harvard University Press, 1998.
5. Bridges, D., Richman, F., *Varieties of Constructive Mathematics*. London Mathematical Society Lecture Note Series, vol. 97. Cambridge: Cambridge University Press, 1987.
6. Brouwer, L.E.J., 'De onbetrouwbaardheid der logische principes', *Tijdschrift voor wijsbegeerte* 2, 152–158, 1908.
7. Brouwer, L.E.J., *Collected Works 1–2*. Heyting (ed.), North-Holland, 1975–76.
8. Burgess, J., *Fixing Frege*. Princeton: Princeton University Press, 2005.
9. Cantor, G., 'Letter to Dedekind', in van Heijenoort (ed.), *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Cambridge, MA: Harvard University Press, 113–117, 1967.
10. Coquand, T., Huet, G., 'The calculus of constructions', *Information and Computation* 76, 95–120, 1988.
11. Curry, H.B., Feys, R., *Combinatory Logic.* Amsterdam, Netherlands: North-Holland, 1958.
12. Dedekind, R., *Was sind und was sollen die Zahlen?* Vieweg, Braunschweig, 1888. English translation 'The Nature and Meaning of Numbers', in R. Dedekind, (ed.), *Essays on the Theory of Numbers*, New York: Dover, 29–115, 1963.
13. Dedekind, R., 'Letter to Keferstein', in van Heijenoort (ed.), *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Cambridge, MA: Harvard University Press, 98–103, 1967.
14. Demopoulos, W. (ed.), *Frege's Philosophy of Mathematics*. Cambridge: Harvard University Press, 1997.
15. Dummett, M., '*The Philosophical Basis of Intuitionistic Logic*', in H.E. Rose and J. Shepherdson (eds.), *Logic Colloquium '73*. Amsterdam: North-Holland, 1975.

16. Dummett, M., *Elements of Intuitionism.* Oxford: Oxford University Press, 1977.
17. Fine, K., *The Limits of Abstraction*. Oxford: Oxford University Press, 2002.
18. Frege, G., *Begriffsschrift: Eine der aritmetischen nachgebildeten Formelsprache des reinen Denkens*. Halle: Louis Nebert, 1879. Translated in van Heijenoort (ed.), *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Cambridge, MA: Harvard University Press, 1967.
19. Frege, G., *Die Grundlagen der Arithmetik: Eine logischmathematische Untersuchung über den Begriff der Zahl*. Breslau: Wilhelm Koebner, 1884. Trans. J. L. Austin. *Foundations of Arithmetic*. Oxford: Blackwell, 1950.
20. Frege, G., *Die Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, 2 vols. Jena: Pohle, 1893/1903. Reprinted 1962. Hildesheim: Olms.
21. Frege, G., *Philosophical and Mathematical Correspondence*, Oxford: Blackwell, 1980.
22. Gentzen, G., 'Die Widerspruchfreihet der Zahlentheorie', *Mathematische Annalen* 112, 493–565, 1936.
23. Goldfarb, W., 'Frege's Conception of Logic', in Floyd and Shieh (eds.), *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*. Oxford: Oxford University Press, 25–41, 2001.
24. Gödel, K., 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38, 173–199, 1931. Translated in van Heijenoort (ed.), *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Cambridge, MA: Harvard University Press, 1967.
25. Gödel, K., 'Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes', *Dialectica* 12, 280–287, 1958.
26. Hale, B., Wright, C., *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon Press, 2001.
27. Heck, R., Jr., 'The development of arithmetic in Frege's Grundgesetze der arithmetic', *Journal of Symbolic Logic* 58, 579–601, 1993. Reprinted in Demopoulos, W. (ed.), *Frege's Philosophy of Mathematics*. Cambridge: Harvard University Press, 1997.
28. Heck, R., Jr., 'The Julius Caesar Objection', in R. Heck (ed.), *Language, Thought, and Logic: Essays in Honour of Michael Dummett*, Oxford: Oxford University Press, 1997.
29. van Heijenoort, J., (ed.) *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Cambridge, MA: Harvard University Press, 1967.
30. Henkin, L., 'Completeness in the theory of types', *Journal of Symbolic Logic* 15, 81–91, 1950.
31. Hesseling, D.E., *Gnomes in the Fog: The Reception of Brouwer's Intuitionism in the 1920s.* Basel: Birkhäuser, 2003.
32. Heyting, A., 'Die Formale Regeln der intuitionistischen Logik', *Sitzungsberichte der Preussische Akademie von Wissenschaften. Physikalischmathematische Klasse*, 42–56, 1930.
33. Hilbert, D., *Grundlagen der Geometrie.* Berlin: Teubner Verlag, 1899.
34. van der Hoeven, G., Moerdijk, I., 'On choice sequences determined by spreads', *Journal of Symbolic Logic*, 49, 908–916, 1984.
35. Howard, W., *The Formulae-as-Types Notion of Construction*', in J.P. Seldin and J.R. Hindley (eds.), *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism.* London: Academic Press, 1980.
36. Kleene, S.C., 'On the interpretation of intuitionistic number theory', *Journal of Symbolic Logic* 10, 109–124, 1945.
37. Kolmogorov, A.N., 'On the principle of excluded middle' (Russian). *Mat. Sb.* 32, 646–667, 1925. Translated in van Heijenoort (ed.), *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Cambridge, MA: Harvard University Press, 414–437, 1967.
38. Lambek, J., Scott, P.J., *Introduction to Higher Order Categorical Logic.* Cambridge: Cambridge University Press, 1986.
39. Lawvere, F.W., 'An elementary theory of the category of sets', *Proceedings of the National Academy of Sciences, USA* 52, 869–872, 1964.

40. Linsky, B., *Russell's Metaphysical Logic*, CSLI Publications, Stanford: Center for the Study of Language and information, 1999.
41. Mancosu, P., *From Brouwer to Hilbert—The Debate on the Foundations of Mathematics in the 1920s*. New York and Oxford: Oxford University Press, 1998.
42. Martin-Löf, P., *Notes on Constructive Mathematics.* Stockholm: Almqvist och Wiksell, 1970.
43. Martin-Löf, P., 'An intuitionistic theory of types', Report, Department of Mathematics, Stockholm University 1972. Reprinted in G. Sambin and J. Smith (eds.), *Twenty-Five Years of Constructive Type Theory,* Oxford: Oxford University Press, 1998.
44. Martin-Löf, P., 'An intuitionistic theory of types: predicative part', in H.E. Rose and J. Shepherdson (eds.) *Logic Colloquium '73.* Amsterdam: North-Holland, 73–118, 1975.
45. Parsons, C., 'Frege's Theory of Number', in Max Black (ed.), *Philosophy in America*. Ithaca: Cornell University Press, 180–203, 1965. Reprinted in Demopoulos, W. (ed.), *Frege's Philosophy of Mathematics*. Cambridge: Harvard University Press, 1997.
46. Peano, G., *Arithmetices Principia, nova methodo exposita*. Bocca, Torino, 1889. English translation in van Heijenoort (ed.), *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Cambridge, MA: Harvard University Press, 83–97, 1967.
47. Reck, E., 'Dedekind's structuralism: an interpretation and partial defense', *Synthese* 137, 369–419, 2003.
48. Shapiro, S., *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford: Oxford University Press, 1991.
49. Sieg, W., 'Hilbert's programs: 1917–1922', *Bulletin of Symbolic Logic*, 5(1), 1–44, 1999.
50. Sieg, W., 'Beyond Hilbert's Reach', in D.B. Malament (ed.), *Reading Natural Philosophy–Essays in the History and Philosophy of Science and Mathematics*. Chicago & La Salle, Illinois: Open Court Press, 363–405, 2002. Reprinted in the present volume.
51. Sieg, W., Schlimm, D., 'Dedekind's analysis of number: systems and axioms', *Synthese* 147(1), 121–170, 2005.
52. Tennant, N., *Anti-Realism and Logic: Truth as Eternal.* Oxford: Oxford University Press, 1987.
53. Troelstra, A.S., *Choice Sequences*. Oxford: Oxford University Press 1977.
54. Troelstra, A.S., van Dalen, D., *Constructivism in Mathematics,* vol. I and II. Amsterdam: North-Holland, 1988.
55. Whitehead, A.N., Russell, B., *Principia Mathematica*, 3 vols, Cambridge: Cambridge University Press, 1910, 1912, 1913. Second edition, 1925 (Vol. 1), 1927 (Vols. 2, 3). Abridged as *Principia Mathematica to *56*, Cambridge: Cambridge University Press, 1962.
56. Wright, C., *Frege's Conception of Numbers as Objects*. Aberdeen: Scots Philosophical Monographs, 1983.
57. Zach, R., 'Hilbert's Program', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2003 Edition, URL = <http://plato.stanford.edu-/archives/fall2003/entries/hilbert-program/>

Part I
Logicism and Neo-Logicism

Protocol Sentences for Lite Logicism

John P. Burgess*

To the memory of Richard Jeffrey

Abstract According to the late Richard Jeffrey's "lite logicism", a kind of a cross between Frege-Russell logicism and Hilbert's formalism, mathematics is logical only in the sense in which physics is empirical: the data of mathematics are logical, as the data of physics are empirical, though in each case a theoretical structure of the science goes far beyond the data. After this view is introduced and compared and contrasted with others, the question just what form the "protocol sentences" or reports of data are to take is examined.

1 Attenuated Empiricism

When the ancient hypothesis of atomism was revived and introduced into modern chemistry by Dalton in order to explain the law of combining volumes, it at first met with a cautious reception. So, too, did Avogadro's further elaboration of atomism, his theory of molecules, and so, too did Arrhenius's yet further elaboration of molecularism, his theory of ions. Yet each of these new hypotheses showed itself capable of predicting phenomena that had not been taken into account when the hypothesis was framed, and in the face of a series of such successful applications, resistance among chemists essentially disappeared before the end of the nineteenth century. Thus Arrhenius, in his Nobel Lecture (1903), after mention of his predecessors back to Democritus and exposition of some of his own principal results, was able to speak as follows:

> I have now described how theories of electrical dissociation have developed from our old ideas about atoms and molecules. We sometimes hear the objection raised, that this viewpoint is perhaps not correct, but only a useful ... working hypothesis. This objection is in

J.P. Burgess (✉)
Department of Philosophy, Princeton University, Princeton, NJ 08544-1006 USA
e-mail: jburgess@princeton.edu

* An earlier version of this paper was presented orally at the August, 2004, Uppsala conference "Logicism, Intuitionism, Formalism: What Has Become of Them?". The present version is a revision of the original in the light of comments by several participants, especially Thierry Coquand and Per Martin-Löf and Göran Sundholm.

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_2, © Springer Science+Business Media B.V. 2009

> fact not an objection at all, for we can never be certain that we have found the ultimate truth. Theories of molecules and atoms are sometimes attacked on philosophical grounds. [But] until a better and more satisfactory theory appears, chemists can continue to use the atomic theory with complete confidence. The position is exactly the same as regards electrolytic dissociation.[1]

While this confident attitude may have been shared by chemists, among some physicists doubts persisted well into the twentieth century. This was partly because it was not easy to achieve a proper understanding of the relationship between phenomenological thermodynamics and statistical mechanics—witness the famous debate between Ludwig Boltzmann and Max Planck's assistant Ernst Zermelo. It was the theoretical work of Einstein (1905) and the experimental work of Perrin (1911) on Brownian motion that proved decisive in bringing the most conservative physicists to share the attitude of the more progressive physicists and of the chemists. By the time of the publication of Perrin's *Atoms* (1913), the scientific debate was closed; any doubts after that time reflect not scientific caution, but philosophical skepticism.

Philosophical debate about the acceptability, and even the intelligibility, of hypotheses about unobservable theoretical posits did continue during the interwar period. Such issues arose in different ways in connection with Russell's logical atomism, with the project of Carnap's *Logische Aufbau der Welt* and more generally with the logical positivism or logical empiricism of the Vienna Circle, with the verificationism of Ayer and the falsificationism of Popper, and with the operationalism and related doctrines of Bridgman and other physicists-turned-philosophers, whose views looked back to those of Mach. In the course of discussions and debates over these matters, a three-level picture of science was elaborated in different ways by different writers, which I would like to begin by reviewing briefly.

Bottom level: protocol sentences. The curious term "protocol sentences" was used for the observation reports at the ground floor or level zero. Underlying such reports are sensory observations, and some thought the observation reports should themselves be formulated in a special language of sensation. But the prevailing view was that they should be formulated in the ordinary language of sensible objects, the everyday language of chairs and tables, knives and forks, and the furniture and utensils of scientific laboratories. So an example of an observation report might be "The beaker is growing warmer," rather than something about heat sense data in the hand holding the beaker.

Observation reports, if true, can always be verified or recognized as true by observation: If the beaker grows warmer it can felt to do so. In this sense, science does not yield any true protocol sentences that observation cannot. What it does is to yield true protocol sentences *more quickly* than observation can: as predictions of observations to come rather than records of observations past.

---

[1] Svante Arrhenius, "Development of the Theory of Electrolytic Dissociation", Nobel Lecture, 11 December, 2003, first published in English in the *Proceedings of the Royal Institution*, 1904, pp. 45–58; now available for downloading in pdf format from nobelprize.org/chemistry/laureates/1903/arrhenius-lecture.pdf. The passage quoted is from the antepenultimate and penultimate paragraphs.

Observation reports can in fact not only be verified if true, but falsified if false: If the beaker remains cool, it can be felt to remain cool. Hence not merely ascriptions of observational predicates to observable objects, but also truth-functional compounds of such ascriptions will be decidable (verifiable if true, falsifiable if false), and may themselves be considered observation reports in an extended sense. Not so with generalizations from observation reports. These take us to the next level.

Middle level: primary laws. The primary or empirical laws at the first floor or level one involve only the observational vocabulary of protocol sentences, being generalizations of observation reports, especially ones in conditional form. An example of a law might be, "In any case, if you take a beaker and pour in first lye, then vinegar, the beaker becomes warmer."

An empirical law can by itself yield observation reports in the extended sense allowing truth-functional compounds, for instance, "In the present case, if this lye and this vinegar are poured into this beaker, it will become warmer;" and an empirical law with some observations reports as auxiliaries can yield other observation reports. Our law together with the observation report, "In the present case, lye and vinegar have been poured into the beaker," will yield the prediction, "In the present case, the beaker will grow warmer". Even one unsuccessful prediction of this kind can refute a universal generalization, while no number successful predictions can prove it; but it is not proof one seeks in science, but only inductive confirmation or corroboration.

While all empirical laws, once conjectured, can be thus tested against observation, only some could in practice be conjectured from an accumulation of observation reports by simple Baconian or enumerative induction from particular to general. Since lye and vinegar are common household substances, the law about them may be an example. Other laws, for instance, those correlating the strength of acids with their electrical conductivity, in practice could hardly have been conjectured except by deducing them from theoretical hypotheses that have themselves been conjectured from an accumulation of other laws. Baconian induction, in practice, does not take us very far in science. Hence we must proceed to the next level.

Top level: theoretical principles. Theoretical principles at the second floor or level two differ from primary laws at level one in that they do not directly yield predictions, even with observation reports as auxiliary hypotheses; that is why empiricist philosophers have tended to be suspicious of them. Least suspect are those that fail to yield predictions only on account of their logical complexity. An example might be, "Whenever a substance that always turns litmus paper blue and a substance that always turns litmus paper red are poured into a beaker, the beaker becomes warmer." This does *not* tell us that if first lye, then vinegar have been poured into a beaker, then the beaker will grow warmer. For to get such a prediction we must first have the laws "Lye always turns litmus paper blue," and "Vinegar always turns litmus paper red." It is only with the help of certain laws and observation reports as auxiliary hypotheses that a logically complex statement can yield observational predictions.

More suspect among empiricists are those theoretical sentences that involve non-observational vocabulary, such as "Whenever a solution with an excess of hydroxide ions and a solution with an excess of hydronium ions are poured into a beaker, the beaker grows warmer." We never get any prediction from a single such theoretical sentence, *even with* the aid of laws and observation reports as auxiliary hypotheses. It always takes several statements in theoretical vocabulary together to yield anything at a lower level. In our example, we would need other sentences involving the theoretical vocabulary of ions, namely, "Lye releases an excess of hydroxide ions in solution" and "Vinegar produces an excess of hydronium ions in solution." It is not theoretical *sentences* that give us predictions, but scientific *theories* consisting of a cluster of such sentences. Though suspect among certain philosophers, theoretical hypotheses, including existential hypotheses positing such unobservable entities as atoms and molecules and ions, have proved indispensable in the practice of science, and that is why scientific debate over the admissibility of such hypotheses has closed.

A philosopher may nonetheless still ask whether what is infeasible in practice, the elimination of reference to unobservable posits, may yet be possible in principle. Could theoretical hypotheses be somehow eliminated? Or at least, could those involving non-observational vocabulary somehow be reduced to logically complex ones that do not? Such a reduction might be possible if theoretical terms like "acid" could be given operational definitions like "substance that always turns litmus paper red," instead of "substance that releases hydronium ions in solution," or something still more theoretical. For a couple of decades after the closure of the scientific debate over atoms, there were philosophers who had hopes that some such reductionist project might succeed. All failed, as was acknowledged in Hempel's 1950 "Problems and Changes" paper, the obituary notice of logical positivism.[2]

So long as there were hopes that reductionism might succeed, a philosopher could hope to be able to combine allegiance to empiricism with allegiance to orthodox physics and chemistry. When reductionism failed, most philosophers who had held such hopes abandoned the strict empiricist view that only notions definable in observational vocabulary are meaningful, in favor of something looser. One fallback position would require, not that each theoretical hypothesis should have a direct observational meaning, but only that each theory, in conjunction with suitable auxiliary hypotheses, should yield some new predictions that can be tested by observation—if not literally, chronologically new, at any rate new in the sense that one did not have these phenomena in mind when framing the theory. Such an attenuated empiricism would not "block the road of inquiry" by ruling out *a priori* hypotheses about unobservable theoretical posits, but it would counsel caution about accepting any such hypothesis so long as all it does is to purport to explain phenomena already known. Such an attenuated empiricism does not seem unscientific; on the contrary, it seems to have been reflected in the initial caution of scientists about atomism and its successive elaborations.

---

[2] Carl G. Hempel, "Problems and Changes in the Empiricist Criterion of Meaning", *Revue Internationale de Philosophie* 41 (1950), pp. 41–63.

Among those philosophers who have chosen to remain faithful to strict empiricism, even if this brings them into conflict with orthodox science, it has become common to feign belief in the reality of atoms while discussing scientific questions and practical issues to which science is relevant. In discussing the greenhouse effect they will talk just as Arrhenius did when he first raised a warning over this issue, and speak of carbon dioxide and other molecules in the atmosphere. But on entering the philosophy room, they will take this all back, and deny any real belief in atoms and molecules. Such skeptical empiricists are likely to meet with the reply that the problem of the reality of atoms, once it is no longer a scientific issue but has become a purely philosophical one, is meaningless. The *locus classicus* for this kind of reply is Rudolf Carnap's "Empiricism, Semantics, and Ontology", which appeared in exactly the middle of the last century (in the same issue of the same journal as Hempel's article, in fact).[3] The appearance of this paper did not mark the end of philosophical debate over atomism, since as Carnap remarks it is in the nature of philosophical debates never to end; but it did end one important phase of that debate.

Since that time the simplistic three-level picture of science I have been describing has itself been subjected to intense criticism, beginning with an attack on the assumption (which no one ever imagined was anything more than an idealization) that scientific vocabulary can be divided into "observational" and "theoretical". But my aim here is not to defend this picture of science, but only to compare it with a certain picture of mathematics, to which I now turn.

2 Attenuated Finitism

Hermann Weyl, in his comments (1927) on Hilbert's second lecture on the foundations of mathematics, presents Hilbert's philosophy of mathematics as being quite consciously modeled on contemporary philosophy of science. He makes the same point, in very similar language, two decades later in his Hilbert obituary notice (1946). In that version he describes Hilbert as pointing to "...the neighboring science of physics where likewise, not the individual statement is verifiable by experiment, but in principle only the system as a whole can be confronted with experience."[4] Elaborating on Weyl's remarks, it may be illuminating to compare in

---

[3] Rudolf Carnap, "Empiricism, Semantics, and Ontology", *Revue Internationale de Philosophie* 41 (1950), pp. 20–40.

[4] Hermann Weyl, "Comments on Hilbert's Second Lecture on the Foundations of Mathematics" (1927), English translation by Stefan Bauer-Mengelberg and Dagfinn Føllesdal, in J. van Heijenoort, ed., *From Frege to Gödel: A Source Book in Mathematical Logic 1879–1931*, pp. 482–484. "David Hilbert and His Mathematical Work", *Bulletin of the American Mathematical Society* 50 (1944), pp. 612–654, reprinted in Constance Reid, *Hilbert*, Berlin: Springer (1970), pp. 245–283. The passage from the latter quoted here is from p. 269 of the reprint, the remark alluded to later about Hilbert being "more papal than the pope" is from the next page. The passages from the former to be quoted at the end of this paper come from its last two paragraphs.

some detail Hilbert's picture of mathematics with picture of science I have just been discussing.

Bottom level: protocol sentences. At level zero in Hilbert's picture come the sentences Hilbert called "*inhaltlich*". This word, as Hilbert used it, means "having content," something not easily expressible in English by a single adjective (hence the frequent use in discussions such as this of the unlovely coinage "contentual"). Such are computation reports, beginning with equations and inequalities between the results of applying to specific numbers such arithmetic operations as addition and multiplication, and beyond these exponentiation and other functions: for instance, "two plus two equals four." Such reports are decidable by computation (verifiable if true, falsifiable if false), and hence so are truth-functional compounds thereof, and beyond these bounded quantificational compounds, which may also be considered computation reports in an extended sense. According to Hilbert, such computation reports are grounded in intuitions (in a quasi-Kantian sense) of syntactic objects, as observation reports in physics or chemistry are grounded in sense-perceptions; but for the moment we may take the bottom-level sentences to be expressed in arithmetical rather than syntactical terms.

Middle level: primary laws. Unbounded universal quantification takes us to the next level, the level of what logicians call $\Pi^0{}_1$ sentences. These are what correspond in Hilbert's picture to empirical laws. As empirical primary laws do not yield empirical protocol sentences beyond those yielded by observation, but rather deliver some of the same ones more quickly, as predictions, so also computational primary laws or $\Pi^0{}_1$ sentences yield no computational protocol sentences not obtainable by calculation, but rather yield computational protocol sentences more quickly. Thus the commutative law for multiplication, which may be expressed by the $\Pi^0{}_1$ sentence $a \cdot b = b \cdot a$ (with initial universal quantifiers, as per convention, left to be tacitly understood) yields

$$123,456,789 \cdot 987,654,321 = 987,654,321 \cdot 123,456,789$$

without our having to undertake the considerable labor of working out the eighteen-digit product $(121,932,631,112,635,269)$. The indirect route through the commutative law is much shorter than the direct route of calculation.

Of course, the proof of the commutative law for multiplication is itself not so very short (since one has to prove the associative and commutative laws for addition, the distributive law, and the associative law for multiplication first). If we count the length of the proof of the commutative law as part of the length of the indirect route (which may not be fair, since after all we only have to prove the law once and we have it available for instantiation by as many specific numbers as we desire), then that route may no longer be shorter than the direct route of computation where nine-digit numbers are involved. But in general, if we compare how long it takes to get a numerical result by reproducing the proof of a $\Pi^0{}_1$ sentence, and then instantiating with specific numbers, with how long it takes simply to do the calculations with those specific numbers, then we find that the length of the first, indirect route grows much more slowly as the specific numbers get larger than does the length of the second, direct route, and for large enough numbers becomes distinctly shorter.

As some empirical laws, Baconian generalizations, are more readily discoverable than others, so also with arithmetical laws expressed by $\Pi^0{}_1$ sentences there is a distinction to be made that is central to Hilbert's philosophy of mathematics. Some arithmetical laws, like the associative and commutative and distributive, are *finistically provable*. They are deducible from the defining equations for the arithmetical functions involved using only a simple form of proof by recursion or mathematical "induction", namely, the *quantifier-free* form. In contrast to arithmetical laws whose usual proofs are of this kind, Hilbert noted that there were others for which the only known deduction uses the infinitistic apparatus of classical mathematics. Such was the status in Hilbert's day of Chebyshev's theorem (a.k.a. Bertrand's postulate, that between any number greater than one and its double there is a prime), which was originally proved using complex analysis.

Top level: theoretical principles. The classical apparatus begins to go beyond the $\Pi^0{}_1$ level already with first-order Peano arithmetic, PA, where we still have only arithmetical *vocabulary*, but sentences of greater logical complexity. When we go still higher, to higher-order arithmetic and Russell's theory of types, PM, or to the axiomatic set theory of Zermelo with the amendment by Frankel, ZFC, we of course have also non-arithmetical vocabulary, the language of classes or sets. All this apparatus Hilbert declared to be "formal" or "ideal" as contrasted with "*inhaltlich*" or "real".

Roughly corresponding to the reductionist projects in philosophy of empirical science was the reductionist project implicit in Hilbert's program, whose aim was to justify classical mathematics to its critics in the following sense. On the plane of philosophical principle, Hilbert would concede that almost the whole of classical mathematics is devoid of computational content, agreeing in this sense even with Paul Gordan, who had criticized Hilbert's own first major result, his famous basis theorem, as being "not mathematics, but theology." On the plane of mathematical practice, however, Hilbert would concede nothing at all, defending even "Cantor's paradise" of transfinite cardinal and ordinal arithmetic, insisting that taking a detour through the merely formal or ideal was in practice indispensable as a route to many results that were themselves real or *inhaltlich*.

Orthodox scientists and mathematicians may say to skeptical critics, "Higher theory, which you declare meaningless, is needed to get certain results that even you consider meaningful, which in practice you cannot obtain." The skeptics will reply, "But how do we know that the meaningful results implied by meaningless higher theories are *true*?" The orthodox may then say, "Any instance can be checked directly (by observation or calculation), and all that have been checked so far that have been found to be true." The critics will reply, "But how do we know that this will continue to be so in the future?" The orthodox may say, "By induction." The skeptic will reply, "And is induction an acceptable method of inference?"

Here a disanalogy emerges between the scientific and the mathematical cases. The orthodox scientist may say, "Induction is used all over science, and must be accepted if one is going to have any science at all." But the orthodox mathematician cannot claim that induction—I mean induction properly so called, not mathematical "induction", better called proof by recursion—is used all over mathematics. On the contrary, in mathematics a requirement of *proof* was the ancient ideal, and after some lapses in the early modern period, the realization of that ideal has been generally demanded since the nineteenth century.

It was this requirement, without analogue in the case of physics or chemistry, that Hilbert aimed to fulfill. To justify to the critic the use of theorems of classical mathematics to derive $\Pi^0{}_1$ sentences and hence computation reports, Hilbert would establish that classical mathematics is *reliable* in the sense that any theorems derived by classical means that are *inhaltlich* are also true. Of course, if such a result is to make any impression on the critic, it must itself be established by methods that are beyond criticism. Hilbert took "finitist" methods in roughly the sense of Kronecker (which apply not only to numerical computation, but also to symbolic computation) to be beyond question; and so his aim was to show *by finitistic reasoning*, that classical mathematics is reliable, a task that quickly reduces to showing *by finitistic reasoning* that classical mathematics is consistent. A finitistic consistency proof for classical mathematics would implicitly provide a method that in principle could be used to convert any classical proof of a sentence, such as Chebyshev's original proof of his theorem, into a (much more tedious and much less perspicuous) finitistic proof. To provide such a consistency proof was the top item on the agenda or program of Hilbert's proof theory or metamathematics.

Gödel's second incompleteness theorem is generally considered to have shown that this program cannot be carried out. Thus strict allegiance to finitism and allegiance to orthodox mathematics cannot be reconciled, any more than strict allegiance to empiricism and to orthodox physics and chemistry. If one wishes to retain orthodox mathematics, one will have to relax or loosen the requirements of finitism, just as one has to relax or loosen the requirements of empiricism if one wishes to retain orthodox chemistry and physics. The mathematical analogue of the loose, attenuated empiricism I described earlier, requiring only that scientific theories should provide new predictions of the results of observation, would be a loose, attenuated finitism requiring only that higher mathematical theories should provide new predictions of the results of computations.

Gödel's second incompleteness theorem in fact shows that this requirement of attenuated finitism is met. As we pass from theories of lower to theories of higher (consistency) strength—from PA to PM, from PM to ZFC, from ZFC to ZFC plus large cardinals—we do get more and more $\Pi^0{}_1$ sentences as theorems. This is because, on the one hand, the consistency of the lower theories can be proved in the higher theories, whereas no theory can prove its own consistency; and on the other hand, the consistency of an axiomatizable theory can be formulated as a $\Pi^0{}_1$ sentence. Gödel's related speed-up theorem tells us that each higher theory, besides yielding $\Pi^0{}_1$ sentences that lower theories did not, also yields more quickly than the lower theories yielded them some $\Pi^0{}_1$ sentences that the lower theories did already yield.

3 Attenuated Feasibilism

Weyl in the 1946 obituary describes Hilbert as being "more papal than the pope" in the sense that Hilbertian finitism was *more* restrictive than Brouwerian intuitionism, the main school of critics of classical mathematics in Hilbert's day. There is no similar language in Weyl's 1927 remarks. Nor did Hilbert, in the lecture on which Weyl was commenting, show much awareness of the distinction between finitism and intuitionism. This lack of awareness is also evident in, for instance, von Neumann's presentation at the 1930 Königsberg symposium on logicism, intuitionism, and formalism, where he replied as representative of Hilbert's formalism to the presentations of Russell's logicism by Carnap and of Brouwer's intuitionism by Heyting.[5]

At Königsberg, von Neumann presented the formalist position as something like a dialectical synthesis of the logicist thesis and its intuitionist antithesis. Because formalism is supposed to preserve something from each of the other schools, he was able to open with polite compliments to the rival schools in phrases like the following:

> Brouwer's sharp formulation of the defects of classical mathematics... Russell's thorough and exact description of its methods (both the good and the bad)...

Such remarks, however, though perhaps appropriate to the occasion, are rather misleading as to the relationship between formalism and the two rival schools.

On the one hand, it is misleading to suggest, as the remark about the "methods" of classical mathematics tends to do, that there is some special relationship between Hilbert's formalism and Russell's PM without mentioning that there is an exactly similar relationship with Zermelo's and Frankel's ZFC. For all that Hilbert's program really required was *some* completely formalizable theory sufficient to encompass the bulk of classical mathematics.

On the other hand (and more importantly in the present context) it is equally misleading to suggest there is a special relationship with Brouwer's intuitionism. In remark about "defects of classical mathematics" just quoted, and in other remarks in von Neumann's exposition, as well as in remarks of Hilbert himself at about this period, we see a tendency to conflate finitistic methods, to which Kronecker had wished to restrict mathematics and to which Hilbert restricted his proof theory or metamathematics, were the same as the intuitionistic methods advocated by Brouwer. It is, of course, now well known that intuitionism is less restrictive than finitism (and ironically the Hilbert basis theorem is an example of a result that admits an intuitionistic but not a finitistic reconstruction).

Yet, had Hilbert and his school recognized at the time the gap between finitism and intuitionism, it might have made no difference. For given that Hilbert's aim

---

[5] J. von Neumann "The Formalist Foundations of Mathematics" (1931), English translation by E. Putnam and G. J. Massey, in P. Benacerraf and H. Putnam, eds., *Philosophy of Mathematics: Selected Readings*, 2nd ed., Cambridge (England): Cambridge University Press (1983), pp. 61–65. The quoted phrases are from the opening paragraph.

concerning foundational questions was to get rid of them once and for all [*einmal aus der Welt zu schaffen*], it arguably was appropriate for him to accept or feign to accept the strongest restrictions any sane critic had advocated, which would have meant finitist and not just intuitionist restrictions, even though intuitionism was at the time the more active movement.

Since Hilbert's time, a still more restrictive philosophy, a *stricter* finitism has emerged, which questions even much of what finitism allows. Finitism admits the operations of addition, multiplication, exponentiation, and more—according to the analysis of Tait, all the functions of Skolem's primitive recursive arithmetic—while stricter finitism balks already at exponentiation. The ground offered for the rejection of exponentiation is that when it is feasible to write down (the decimal or binary numerals for) two numbers, it is feasible to compute their sum and product, but not, in general, the value of the first raised to the power of the second: If $a = 123,456,789$ and $b = 987,654,321$, we can with some strain compute $a \cdot b$, but a^b is beyond us. A further, related restriction is that while truth-functional compounds of equations and inequalities involving addition and multiplication are allowed, bounded quantifiers are banned, also on the grounds of infeasibility. In short, where finitism insists on computability, stricter finitism insists on feasibility, and therefore might well be called "feasibilism".

It is natural to feel some suspicion of a position based on such a vague notion as "feasibility", but my aim here is not to defend stricter finitism in any stronger sense than that of maintaining that it is not insane. What I want to urge is that if Hilbert had known or conceived of the feasibilist position, it would have been appropriate for him, given his aims, to have imposed not just finitist but feasibilist restrictions in his program. Had he done so, would the Gödelian conclusions about the Hilbert program still hold? As higher scientific theories do some observational work, making new predictions of the results of possible observations, so do higher mathematical theories do some computational work, making new predictions of the results of possible calculations. But is this still true if we restrict our attention to calculations that are not just possible in principle, but feasible in practice? That is to say, if we now restrict our attention to those ${\Pi^0}_1$ sentences admitted by feasibilists, is it still true that higher theories give new ${\Pi^0}_1$ sentences of this restricted kind?

In connection with such questions it will be useful to introduce a bit of jargon. Let us call a class of ${\Pi^0}_1$ sentences *revelatory* if stronger and stronger theories do prove more and more of them. Thus Gödel's work tells us that the full class of *all* ${\Pi^0}_1$ sentences is revelatory, and our question now is whether the restricted subclass consisting of universal generalizations of truth-functional compounds of equations and inequalities involving addition and multiplication is revelatory. The question is non-trivial, but an affirmative answer emerges as a corollary to the proof of a theorem due to Yuri Matiyasevich, building on earlier work of Martin Davis, Hilary Putnam, and Julia Robinson: the solution to the tenth of the problems on Hilbert's famous list of 1900.

Let me begin with what we get from this work. We may define a *DPR sentence* to be (the universal closure of) a formula expressing the inequality of two exponential polynomials. An example is provided by Fermat's conjecture (a.k.a. Wiles's theorem). That result would be most naturally written as a conditional, thus:

$$\mathbf{1} \leq a \;\&\; \mathbf{1} \leq b \;\&\; \mathbf{1} \leq c \;\&\; \mathbf{3} \leq d \rightarrow a^d + b^d \neq c^d$$

But it can be written equivalently as a DPR sentence, thus:

$$(a + \mathbf{1})^{(d+\mathbf{3})} + (b + \mathbf{1})^{(d+\mathbf{3})} \neq (c + \mathbf{1})^{(d+\mathbf{3})}$$

The work of Davis, Putnam, and Robinson in effect reduces $\Pi^0{}_1$ sentences involving arbitrary primitive recursive functions to DPR sentences, which are $\Pi^0{}_1$ sentences involving only addition, multiplication, and exponentiation. To give a very simple example, consider an *identity* between exponential polynomials, such as the first law of exponents:

$$a^{(b+c)} = a^b \cdot a^c$$

This is equivalent to a pair of DPR sentences, one telling us the right-hand side of the identity is not less than the left-hand side, and the other telling us the reverse, thus:

$$a^{(b+c)} \neq a^b \cdot a^c + d + \mathbf{1} \qquad\qquad a^{(b+c)} + d + \mathbf{1} \neq a^b \cdot a^c$$

The Davis-Putnam-Robinson theorem tells us something like this reduction to DPR formulas can be achieved for any $\Pi^0{}_1$ formula, even ones involving superexponentiation and higher primitive recursive functions. A corollary to the proof is that the class of DPR sentences is revelatory.

Turning now to Matiyasevich's theorem, let use define *Matiyasevich sentences* the same way we defined DPR sentences, but with ordinary polynomials in place of exponential polynomials. Then Matiyasevich, by proving a certain key conjecture of Robinson, establishes for Matiyasevich sentences what she had established for DPR sentences. It follows that the class of Matiyasevich sentences is revelatory. Since these form a subclass of the class of universal closures of truth-functional compounds of equations and inequalities involving addition and multiplication, the latter class, which is the class of $\Pi^0{}_1$ sentences admitted by feasibilists, is *a fortiori* revelatory.

And this is as far as we can go. For it is a corollary to the proof of Presburger's theorem on the decidability of arithmetic with addition alone that all true $\Pi^0{}_1$ sentences involving addition alone are finitistically provable. Moving on to higher theories will not give any new results about addition alone. But an ultrastrict finitism or ultrafeasibilism that disallowed even multiplication would be insane, and so need not concern us.

4 What of Applications?

A point Weyl fails to mention is that Hilbert's picture of mathematics is essentially a picture of *pure* mathematics only, saying little directly about applications. Formalism was not alone in essentially ignoring applications, for intuitionism did the same. In the case of intuitionism a possible explanation is that Brouwer's attitude towards applied mathematics was essentially hostile. The attitude of Hilbert and associates was radically different, and so we must seek some alternative explanation for *their* philosophical inattention to applications. Perhaps such an alternative explanation can be found in the hypothesis that Hilbert is merely following Brouwer, aiming to reply to him on his own ground. Perhaps such an alternative explanation can be found in the fact that the applications that most concerned Hilbert and von Neumann and Weyl were applications to advanced and sophisticated physical theories—quantum theory and relativity theory—that were considered by the prevailing empiricist philosophy of science of the period to be themselves "ideal" rather than *inhaltlich*. But let me leave this question to the historians.

Mathematics, of course, has other applications besides its applications to sophisticated physical theories, and an audience of logicians will be well aware that higher mathematics has applications to first-order logic. To be sure, if one uses mathematics to establish the validity of some first-order schema, then one could also in principle have established it instead using the kind of proof-procedure that one finds in textbooks (or that are used in automated theorem-proving programs). That is what the Gödel completeness theorem tells us. Nonetheless, an indirect route to a first-order schema using mathematical methods can sometimes be much quicker than a direct route using only textbook methods.

For one instance, working in PM or higher-order logic, or even just working in second-order logic, one can sometimes obtain a first-order logical schema much more quickly than one could working in first-order logic alone. The best known examples of this phenomenon may be certain "curious inferences" published the late George Boolos.[6] For another instance, working in ZFC, one can sometimes derive syntactic conclusions from semantic considerations; that is, one can sometimes establish set-theoretically that a first-order schema holds in all models more quickly than one can construct a proof of it. The best known examples may be ones pointed out long ago by Alfred Tarski: the completeness of the theory of real closed fields allows one to conclude, when an algebraic theorem has been proved for the real field by analytic methods, that the theorem in fact holds for all real closed fields, even if one has no idea how to deduce the theorem from the axioms for such fields. But there is no mystery why Hilbert and associates did not consider these kinds of applications of mathematics to logic, since they were only discovered at a later period.

---

[6] George Boolos, "A Curious Inference", *Journal of Philosophical Logic* 16 (1987), pp. 1–12, reprinted in George Boolos, *Logic, Logic, and Logic*, Cambridge (Massachusetts): Harvard University Press (1998), pp. 376–382.

Mathematics, however, has yet other applications, neither sophisticated nor historically recent, beginning with the primæval use of arithmetical facts or laws, according which, from something like the premise that "two plus two equals four" one may infer something like the conclusion that "if exactly two sheep jump the fence in the morning, and exactly two sheep jump the fence in the evening, and no sheep jumps the fence both morning and evening, then exactly four sheep jump the fence either morning or evening." (This kind of basic application of *cardinal* concepts is what will be of direct concern here, but it should be mentioned as a topic for future consideration that the question of basic applications of *ordinal* concepts also invites examination.) Hilbert gives a very sketchy account of such applications, but does not elaborate and does not even seem to attach too much importance to the matter. This is somewhat surprising, especially since the logicists made so much of these applications. Roughly speaking, the logicist analysis makes an arithmetical fact or law like "two plus two equals four" essentially a generalization whose instances are conclusions like the one about the sheep.

Such applications can be viewed as applications to first-order logic, at least if we expand our notion of first-order logic to take in numerically definite quantifiers. The existential quantifier $\exists$ of first-order logic means "there exists at least one". We can introduce as an abbreviation the quantifier $\exists!$ meaning "there exists exactly one" in the usual way. But we can also introduce as abbreviations the quantifiers $\exists_{\mathbf{2}}$, $\exists_{\mathbf{3}}$, $\exists_{\mathbf{4}}$, and so on, meaning "there exist at least two" and "there exist at least three" and "there exist at least four" and so on. For example, $\exists_{\mathbf{17}} x Ax$ may be taken to abbreviate the following:

$$\forall x_1 \forall x_2 \ldots \forall x_{16} \exists x (Ax \;\&\; x \neq x_1 \;\&\; x \neq x_2 \;\&\; \ldots \;\&\; x \neq x_{16})$$

(There are also other possibilities.) Then "there exist more than one" and "there exist more than two" and so on are equivalent to "there exist at least two" and "there exist at least three" and so on; and "fewer than" is the negation of "not fewer than" or "at least", while "at most" or "not more than" is the negation of "more than", so we can obtain such quantifiers also. Further, "exactly" amounts to the conjunction "not more and not fewer" or "at most and at least", so we also get the quantifiers $\exists_{\mathbf{2}}!$, $\exists_{\mathbf{3}}!$, $\exists_{\mathbf{4}}!$, and so on, meaning "there exist exactly two" and "there exist exactly three" and "there exist exactly four."

I will henceforth understand "first-order" logic to admit such numerically definite quantifiers. Admitting such quantifiers amounts to admitting the use of numerals as *adjectives*, as in "There are three cups on the table," in contrast to their use as *nouns*, as in "Three is a prime number." The adjectival use goes back to prehistoric times, while the introduction of the nominal use is a roughly datable historical event.

On a logicist analysis, there is an absolutely immediate connection between laws of cardinal arithmetic and certain second-order logical truths, and an almost immediate connection between laws of finite arithmetic and the truth of certain *universal* second-order logical truths, or what comes to the same thing, the validity of certain first-order schemata. (A first-order schema is in effect a universal second-order sentence with the initial universal second-order quantifiers left to be tacitly understood.)

In particular, "two plus two equals four" is in effect analyzed as meaning something like the following generalization, of which the example about the sheep above is an instance:

$$\exists_2!xAx \;\&\; \exists_2!xBx \;\&\; \sim \exists x(Ax \;\&\; Bx) \rightarrow \exists_4!x(Ax \vee Bx) \qquad (1)$$

Now no Hilbertian will accept a logicist analysis of arithmetical facts or laws. Hilbert and his most orthodox followers would insist that the arithmetic is founded on syntactic intuition, not logical deduction; some might even claim that syntactic intuition is needed simple to follow logical deductions. But setting aside all questions of *analysis*, the Hilbertian picture of mathematics must be considered radically incomplete until it adds some provision—I mean, some *formal* provision—for logical *applications*. In short, Hilbert's picture needs to be supplemented with some account of how first-order schemata like (1) can be obtained from arithmetical facts or laws. I will be proposing such a supplement below.

One might imagine the route from a $\Pi^0{}_1$ sentence to a variety of first-order schemata proceeding in either of two ways. On one approach, various substitutions of particular numerals $\mathbf{0}$, $\mathbf{1}$, $\mathbf{2}, \ldots$ for variables in the $\Pi^0{}_1$ sentence will produce a variety of numerical instances, and each of these would be associated with some first-order schema. On another approach, the $\Pi^0{}_1$ sentence would be associated with a single non-first-order schema involving symbols $\exists_u, \exists_v, \exists_w, \ldots$ with variable subscripts, and then various substitutions of particular numerals $\mathbf{0}$, $\mathbf{1}$, $\mathbf{2}, \ldots$ for these variables would produce a variety of first-order schemata. The two approaches are in some sense equivalent, and the choice between them a matter of taste. My proposal below will be formulated the first way, but proposals formulated the second way can also be found in the literature.

5 Lite Logicism

In particular, several such proposals have been put forward by the late Richard Jeffrey, in connection with the perspective on philosophy of mathematics that he called "lite logicism" or "logicism lite", to which I now turn. In American commercial English, "lite" is a deliberate misspelling of "light" used in marketing low-fat and low-calorie food and beverages. For instance, lite "beer" is a fluid with approximately the calorie content and approximately the taste of a fifty–fifty mixture of real beer and soda water. Jeffrey's lite "logicism" may be described as a fifty–fifty mixture of real logicism and formalism.

Lite logicism differs from traditional Frege-Russell logicism or contemporary Wright-Hale neo-logicism in that it makes no attempt to reduce all of mathematics to logic. Lite logicism differs from formalism in that it replaces what I will call the *Hilbert proportion*, namely, the following:

computational : mathematics :: empirical : physics by the following *Jeffrey proportion*:

logical : mathematics :: empirical : physics

On either view, just as orthodox physics is "empirical" only in the attenuated sense that its *data* are reports of empirical observations, so orthodox mathematics is "computational" or "logical" as the case may be only in a similarly attenuated sense. Lite logicism agrees with formalism that mathematics is different from physics or chemistry because it starts from data of a different kind, and similar to natural science because it builds on these data an elaborate theoretical apparatus to bring out patterns and regularities among them. Lite logicism differs from formalism over the nature of the data, taking these to be not computational but rather logical, thus far agreeing with real logicism in assigning great importance to the connection between arithmetic and logic. Lite logicism seeks to picture mathematics as a theoretical structure built upon and testable against, but not reducible to, logical data like the validity of certain first-order schemata. It is these schemata, rather than computation reports, that play the role of protocol sentences in the lite logicist picture of mathematics.

As I have said, Jeffrey made *several* proposals as to how arithmetical facts or laws are to be associated with logical schemata.[7] It would be impossible to argue with any serious degree of informal rigor that any such proposal is optimal without first undertaking some examination of what the desiderata are when one undertakes to extract logical schemata from arithmetical laws and facts; and such an examination was more than Jeffrey had time for in his last years. In taking up the question where he left it, I will begin with such an examination, albeit a rather cursory one, bringing out just two points.

Supposing we are to instantiate $\Pi^0{}_1$ sentences with numerals and then translate the results into logical schemata, the first point to be noted is that from a lite logicist point of view it is not a requirement, and not really even a desideratum, that *every* $\Pi^0{}_1$ sentence should yield logical schemata. For the lite logicist has abandoned the ambition of reducing mathematics to logic. A second point is that it *should* be a desideratum or even a requirement, even for a lite or diluted or attenuated logicism, that stronger and stronger mathematical theories should give us more and more in the way of new quick routes to logical schemata. At any rate, such a requirement would be the obvious analogue of what attenuated finitism and attenuated empiricism require. Let me elaborate a bit on this desideratum or requirement.

It will turn out on the proposal I will eventually be making that the following $\Pi^0{}_1$ sentence:

[7] Two published versions are given in "Logicism 2000", in S. Stich and A. Morton, eds., *Benaceraff and His Critics*, London: Blackwell (1996), pp. 160–164, and in his last published paper, "Logicism Lite", *Philosophy of Science* 69 (2002), pp. 447–451. The latter treatment also cites work of Fernando Ferreira that, though not intended as a contribution to Jeffrey's own project, could be adapted to that purpose, so there are actually three versions.

$$\sim (c \leq a \;\&\; b < c \;\&\; a \leq b) \tag{2}$$

will by instantiation with specific numerals followed up with translation from arithmetic to logic yield the *pigeon-hole* first-order schemata $P_{\mathbf{0}},\ P_{\mathbf{1}},\ P_{\mathbf{2}}$, and so on. Here in general $P_{\mathbf{m}}$, for any numeral $\mathbf{m} = \mathbf{0}$ or $\mathbf{1}$ or $\mathbf{2}$ or whatever, says the following:

$$\sim (\exists_{\mathbf{m}} x Ax \;\&\; \sim \exists_{\mathbf{m}} x Bx \;\&\; (\forall x(Ax \to \exists y(By \;\&\; Rxy)) \;\&\;$$
$$\forall x \forall y \forall z(Ax \;\&\; Ay \;\&\; Bz \;\&\; Rxz \;\&\; Ryz \to x = y)))$$

or what is trivially equivalent and more perspicuous, the following:

$$\exists_{\mathbf{m}} x Ax \;\&\; \sim \exists_{\mathbf{m}} x Bx \;\&\; \forall x(Ax \to \exists y(By \;\&\; Rxy)) \to$$
$$\exists x \exists y \exists z(Ax \;\&\; Ay \;\&\; Bz \;\&\; Rxz \;\&\; Ryz \;\&\; x \neq y))$$

For say $\mathbf{m} = \mathbf{17}$, this is the first-order schema one of whose instances says that "if you have at least seventeen pigeons and fewer than seventeen pigeon-holes, and if every pigeon goes into a hole, then some hole will have more than one pigeon in it."

There is a substantial literature on the lengths of the shortest proofs of pigeon-hole principles under different formalizations of the principles different types of proof-procedures, some of which do produce proofs quite quickly.[8] What I want to say here is that *even so* obtaining the $P_{\mathbf{m}}$ from (2) by instantiation and translation is *even quicker*; indeed, it is essentially instantaneous, requiring only as much time as it takes to write $P_{\mathbf{m}}$ down. At any rate, we certainly have here a *new* quick way of obtaining $P_{\mathbf{m}}$. So we get something new by applying arithmetic to logic, and mathematics is doing some logical work.

But the part of mathematics doing the work in this particular example is the finitist part. For the law (2) involved in this example is obtainable already in very weak systems of arithmetic. What I maintained above was that the attenuated logicist ought to require that further strengthenings beyond weak systems of arithmetic to PA and PM and ZFC and beyond should do more logical work, in the sense that it should continue to give us new quick ways of obtaining first-order logical schemata.

To meet this requirement we need a method of obtaining logical schemata that will apply to enough $\Pi^0{}_1$ sentences even if not to all, where "enough" has the precise meaning of "enough to constitute a revelatory class" in the sense introduced in the earlier discussion of Davis, Putnam, Robinson, and Matiyasevich. What I will be proposing, therefore, is a way of obtaining logical schemata from arithmetical laws and facts that applies to a class of $\Pi^0{}_1$ sentences including all Matiyasevich sentences.

[8] The best known results are due to Sam Buss, according to whom (personal communication) proofs may be obtain of lengths that are not merely polynomial functions of the length of the principle to be proved, but polynomial functions of very low degree, quadratic or thereabouts.

6 Protocol Sentences

Let me without further preliminaries begin to present the proposed translation from arithmetic to logic, whose products will play for lite logicism the role corresponding to observation reports for empiricism or computation reports for formalism—the role of protocol sentences. Consider a $\Pi^0{}_1$ sentence such as the following:

$$\mathbf{1} \leq a \;\&\; \mathbf{1} \leq b \rightarrow (a+b)\cdot(a+b) \neq (a\cdot a)+(b\cdot b) \tag{3}$$

Such sentences can be reduced to Matiyasevich sentences, and in the case of (3) it is easy to see what Matiyasevich sentence we get:

$$((a+\mathbf{1})+(b+\mathbf{1}))\cdot((a+\mathbf{1})+(b+\mathbf{1})) \neq ((a+\mathbf{1})\cdot(a+\mathbf{1}))+((b+\mathbf{1})\cdot(b+\mathbf{1})) \tag{4}$$

By standard logician's tricks, introducing auxiliary variables, (4) will be equivalent to the following:

$$\begin{aligned} \sim(c = \mathbf{1} \;\&\; a+c = d \;\&\; b+c = e \;\&\; d+e = f \;\&\; f\cdot f = g \;\&\\ d\cdot d = h \;\&\; e\cdot e = i \;\&\; h+i = j \;\&\; g = j) \end{aligned} \tag{5}$$

(I mean, of course, that the *universal closures* of (3, 4, 5) are all equivalent.) Let us call $\Pi^0{}_1$ sentences obtained in this way *reduced* Matiyasevich sentences. Any specific numerical instance will be equivalent to a variant with a couple more conjuncts. For example, the instance of (5) with the numeral **3** substituted for the variable a and **2** substituted for b will be equivalent to the following:

$$\begin{aligned} \sim(a = \mathbf{3} \;\&\; b = \mathbf{2} \;\&\; c = \mathbf{1} \;\&\; a+c = d \;\&\; b+c = e \;\&\\ d+e = f \;\&\; f\cdot f = g \;\&\; d\cdot d = h \;\&\; e\cdot e = i \;\&\; h+i = j \;\&\; g = j) \end{aligned}$$

Let us call sentences obtained in this way *specialized* Matiyasevich sentences.

What I will do will be to introduce a class $\mathcal{R}$ properly including the reduced Matiyasevich sentences, a subclass $\mathcal{S}$ thereof properly including the specialized Matiyasevich sentences, and a method of translating any arithmetical sentence ϕ in the class $\mathcal{R}$ into a first-order schema ϕ^* in such a way that if ϕ is in the subclass $\mathcal{S}$ then ϕ^* is logically valid. (For ϕ in $\mathcal{R}$ but not in $\mathcal{S}$ in general ϕ^* will only be valid over finite domains.) The sentences in $\mathcal{R}$ will be negated conjunctions. What I will describe directly is how to translate a sentence ϕ in $\mathcal{R}$ into negated conjunction $\phi^\dagger$ with one conjunct $\psi^\dagger$ for each conjunct ψ of the original ϕ. The $\psi^\dagger$ will be in some cases first-order, and in other cases existential second-order. But the existential quantifiers can all be brought out in front of the conjunction, and then changing negated existential quantification to universally quantified negation we obtain a universal second-order sentence, which when its initial universal second-order quantifiers are left tacit amounts to a first-order schema ϕ^*.

To specify the class $\mathcal{R}$, it will suffice to specify what sorts of formulas may appear as conjuncts in the negated conjunction. We may allow the following types, wherein *a, b, c* may be any variables, and **m** any specific numeral:

(6)	$a = \mathbf{m}$	(7)	$a \leq \mathbf{m}$	(8)	$a < \mathbf{m}$
(9)	$a \neq \mathbf{m}$	(10)	$a \geq \mathbf{m}$	(11)	$a > \mathbf{m}$
(12)	$a = b$	(13)	$a \leq b$	(14)	$a < b$
(15)	$a \neq b$	(16)	$a \geq b$	(17)	$a > b$
(18)	$a = b + c$	(19)	$a = b \cdot c$		

The specification of the subclass $\mathcal{S}$ will require some preliminaries. Given ϕ in $\mathcal{R}$, let us define recursively or inductively what it is for a variable in ϕ to be *capped*. First, variables in (any conjuncts of types) (6, 7, 8) are capped. Further, if either variable in (12) is capped, so is the other; if the variable on the right-hand side in (13) or (14) is capped, so is the variable on the left-hand side; if the variable on the left-hand side in (16) or (17) is capped, so is the variable on the right-hand side. Finally, if the variables on the right-hand side in (18) or (19) are capped, so is the variable on the left-hand side. Intuitively, if a is capped, then the conjunction implies an inequality of the form $a \leq \mathbf{n}$, where **n** is some specific numeral. A little thought shows that as all reduced Matiyasevich sentences are in $\mathcal{R}$, so all specialized Matiyasevich sentences are in $\mathcal{S}$.

In the translation from arithmetic to logic each variable *a, b, c* will be replaced by a one-place predicate *A, B, C*. In describing how an arithmetical conjunct ψ is to be translated into a first-order or existential second-order $\psi^\dagger$, I will discuss only the five types (7), (13), (14), (18) and (19), leaving the others, which are readily reducible to these, as exercises. Intuitively, $\psi^\dagger$ will say that ψ holds if *a, b, c* are the numbers of x such that *Ax, Bx, Cx*. Thus for (7) the translation is $\exists_{\mathbf{m}} x Ax$. For (13) the translation begins with $\exists R$, after which comes the following:

$$\forall x(Ax \rightarrow \exists! y(By \,\&\, Rxy)) \tag{20}$$

For (14) in place of (13) we have the following in place of (20):

$$\exists z(Bz \,\&\, \forall x(Ax \rightarrow \exists! y(By \,\&\, y \neq z \,\&\, Rxy))) \tag{21}$$

Note that (21) expresses that the number of x such that Ax is less than the number of x such that Bx only provided that the latter number is finite, which will be guaranteed if our original ϕ is in $\mathcal{S}$ (but not in general if ϕ is only in $\mathcal{R}$).

As (20) says that R maps the x such that Ax one-to-one into the y such that By, so the item corresponding to (20) for (18) in place of (13) will say that R maps the y such that By one-to-one onto some of the x such that Ax and maps the z such that Cz one-to-one onto the rest of the x such that Ax. Something like $\mathbf{2} + \mathbf{2} = \mathbf{4}$ can be expressed by something like the following sentence ϕ in $\mathcal{S}$:

$$\sim (a = \mathbf{2} \,\&\, b = \mathbf{2} \,\&\, c = a + b \,\&\, c \neq \mathbf{4})$$

When one works out ϕ^*, one will find that it is *not* identical with (1) above, but rather something logically stronger that logically implies (1). Turning from addition to multiplication, the item corresponding to (20) for (19) in place of (13) will involve a three-place rather than a two-place auxiliary predicate R. What it will say is that R maps the pairs consisting of an x such that Ax and a y such that By one-to-one onto the z such that Cz.

The citation of Matiyasevich's theorem is not really needed: the Davis-Putnam-Robinson theorem would do, since there is an existential second-order sentence that expresses that the number of x such that Ax equals the number of y such that By raised to the power of the number of z such that Cz, provided this last number is finite. In this case what is said about the three-place auxiliary predicate R is the conjunction of the following:

For each x such that Ax, the relation $Syz \leftrightarrow Rxyz$ defines a function s_x from the z such that Cz to the y such that By.

For distinct w and x such that Aw and Ax, the associated functions s_w and s_x do not give the same value for every argument.

For any x and y and z such that Ax and By and Cz there is a w such that Aw with the property that $s_w(z) = y$ and the values of s_w and s_x are the same for all other arguments.

Indeed, even less is needed, since even more operations beyond exponentiation can be expressed in the same way and sense. And with this remark I conclude my discussion of protocol sentences for lite logicism.

7 Instrumentalist Philosophies

What I have emphasized is that the class of protocol sentences is broad enough that we join the attenuated logicists in viewing mathematics simply as an engine or instrument for deriving logical results in the form of such protocol sentences, then we find that higher and higher mathematical theories do more and more logical work, just as higher and higher scientific theories do more empirical work. But is such an attenuated logicism, or the analogous attenuated empiricism that views science simply as an engine or instrument for deriving observational predictions, a tenable philosophical position? Is it really plausible to view theoretical posits such as numbers and sets or classes in mathematics, or atoms and molecules and ions in chemistry and physics as mere "useful fictions"? Such instrumentalist—or as they are now often called, "fictionalist"—positions do appeal to many skeptical philosophers. It must be acknowledged, however, that they hardly accord with the attitudes of working scientists and mathematicians.

This, too, is something Weyl acknowledged and emphasized long ago. Of physics he spoke in the 1927 "Comments" as follows:

> It has been said that physics is concerned only with establishing pointer coincidences. . . . But, if we are honest, we must admit that our theoretical interest does not attach exclu-

> sively or even primarily to the "real propositions", the report that this pointer coincides with that part of the scale; it attaches, rather, to the ideal assumptions that *according to the theory* disclose themselves in such coincidences, but of which no perception gives the full meaning—as, for example, the assumption of the electron as a universal elementary quantum of electricity.

And what Weyl said about Machian empiricist philosophies of physics was intended to apply also to Hilbertian formalist philosophies of mathematics—and applies as well to Jeffrey's cross between logicism and formalism, for which I have been striving here to provide only an adequate technical formulation, not any sort of philosophical defense. Is the attitude of the skeptic looking at mathematics from the outside and inclining towards this or some other form of "fictionalism" better justified than the very different attitude of the working mathematician, whose theoretical interest lies in, say, the Riemann hypothesis rather than in any first-order schemata or computation reports or whatever that the Riemann hypothesis, if proved, might help us derive? Here I can only agree with Weyl:

> What "truth" or objectivity can be ascribed to this theoretic construction of the world, which presses far beyond the given, is a profound philosophical question.

Frege's Context Principle and Reference to Natural Numbers

Øystein Linnebo

Abstract Frege proposed that his Context Principle—which says that a word has meaning only in the context of a proposition—can be used to explain reference, both in general and to mathematical objects in particular. I develop a version of this proposal and outline answers to some important challenges that the resulting account of reference faces. Then I show how this account can be applied to arithmetic to yield an explanation of our reference to the natural numbers and of their metaphysical status.

Grundlagen [7] Section 62 Frege raises a question that has dominated much of recent philosophy of mathematics: 'How, then, are the numbers to be given to us, if we cannot have any ideas or intuitions of them?' The problem is of course that numbers, unlike tables and chairs, cannot be perceived; nor can they be observed with the help of modern technology, as electrons and DNA molecules can. How can we then refer to numbers and other kinds of abstract objects, let alone gain knowledge of them?

It is important to get clear on the nature of this question. What the question calls for is not an account of *what objects* various numerals (or their mental counterparts) refer to. Giving such an account is easy; for instance, '7' and 'VII' refer to 7. Rather, what the question calls for is an account of what facts about reference *consist in*. For instance, we would like to know how it comes about that '7' and 'VII' refer to 7. This question is a perfectly reasonable one. For the fact that a term or representation manages to refer to an object external to itself can hardly be a primitive fact but must have some explanation.

Frege's next sentence makes a proposal about how this question can be addressed. 'Since it is only in the context of a sentence that words have any meaning, our problem becomes this: To define the sense of a sentence in which a number word occurs.' The doctrine that words have meaning only in the context of a sentence has

Ø. Linnebo (✉)
Department of Philosophy, University of Bristol, Bristol, UK
e-mail: Oystein.Linnebo@bristol.ac.uk

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341, DOI 10.1007/978-1-4020-8926-8_3, © Springer Science+Business Media B.V. 2009

become known as the *Context Principle*.[1] What Frege proposes is that the Context Principle has an essential role to play in the explanation of reference, both in general and to numbers and other abstract objects in particular. I will refer to this as *Frege's Proposal*. The idea is to translate the problem of explaining how a singular term comes to refer into the problem of explaining how certain complete sentences involving this term come to be meaningful.

In this paper I develop an account of reference based on Frege's Proposal (Section 1) and outline answers to some important challenges that this account faces (Section 2). Then I show how this account can be applied to arithmetic to yield an explanation of our reference to the natural numbers (Section 3) and of their metaphysical status (Section 4). Given the daunting magnitude and difficulty of the questions I discuss, all I can reasonably hope to do in this paper is to outline a research programme and to begin exploring its strengths and possible weaknesses. A more complete treatment of the questions I discuss would no doubt require an entire book. I should also mention that my present goal is systematic, not exegetical. Although I believe Frege anticipated many aspects of the account that I develop, I do not claim that he would have agreed with all of it.

1 Towards a Fregean Account of Reference

Assume we want to explain how singular terms of a certain kind come to refer. According to Frege's Proposal, we can do this by explaining how complete sentences involving this kind of singular terms come to be meaningful. Before we attempt to provide any such explanations, it will be useful to simplify the problem somewhat.

The first simplification is to focus on thought rather than on language. Our modified *explanandum* is then what is involved in someone's capacity for singular reference to various sorts of object.[2] The proposal is that an adequate explanation of this can be given by explaining what is involved in the person's capacity for understanding complete thoughts concerning objects of the sort in question.[3] This simplification will allow us to concentrate on an individual person rather than on a whole language community. This is a huge simplification. For instance, if reference involves a notion of Fregean sense, then this sense will now be allowed to vary with

[1] See also ibid. pp. x, 71, and 116. I have changed the translation of 'Satz' from 'proposition' to 'sentence'. This is reasonable, given that Frege talks about *words* occurring in a *Satz*.

[2] In the terminology of [5], my goal is to explain what our understanding of the relevant kind of 'fundamental Ideas' consists in.

[3] Strictly speaking, I here collapse two steps. The first step is Frege's suggestion that questions concerning singular reference be addressed in terms of analogous questions concerning complete thoughts. In particular, in virtue of what does a physical state of an agent have a particular thought as its content? The second step is to approach this question about thoughts in terms of the notion of understanding. Doing so is quite natural; for in order to stand in some propositional attitude to a thought, one presumably needs to *understand* that thought.

each individual act of reference. In contrast, if a notion of sense is to be attached to an expression of a public language, then this sense will have to be shared by every competent speaker of this language. (For the purposes of this paper I will not have anything to say about how linguistic expressions come to refer.)

The second simplification is to focus on *canonical* cases of singular reference. These are certain maximally direct ways of referring to objects, where the referent is 'directly present' to the thinker. For instance, referring to a person whom I see immediately in front of me is canonical, whereas referring to Napoleon is not. More examples of canonical reference will be presented shortly.[4]

The third simplification is to begin by explaining someone's understanding of identity statements before attempting to explain his understanding of thoughts more generally. This strategy is adopted by Frege himself in *Grundlagen*.[5] The rationale is that, before one can understand what it means for an object to possess properties and stand in relations, one needs to know how to distinguish the object from other objects and how to re-identify it when presented with it in alternative ways.[6]

With these three modifications, Frege's Proposal becomes the following: We can explain what is involved in someone's capacity for canonical singular reference to objects of a certain kind by explaining what is involved in his or her capacity for understanding identities concerning such objects. This would translate the problem of explaining our capacity for singular reference into the related but different problem of explaining our capacity for understanding identity statements.

Frege suggests an ingenious way in which his proposal can be carried out. The core idea is that canonical reference has a rich and systematic structure. Firstly, objects are always presented to us only via some of their parts or aspects. And secondly, we have a grasp of how two such parts or aspects must be related for them to pick out the same object. Here are some examples.[7]

(a) *Physical bodies*. A physical body is most directly presented in perception, where we causally interact with one or more of its spatiotemporal parts. Two such parts determine the same physical body just in case they are connected through a continuous stretch of solid stuff, all of which belongs to a common unit of motion.[8]

---

[4] Following Michael Dummett and Gareth Evans I believe non-canonical reference must be explained in terms of someone's ability to recognize the referent when presented with it in a canonical way. See [3], pp. 231–239 and [5], pp. 109–112. (In a more complete treatment, I would attempt to give a more explicit account of non-canonical reference.)

[5] Frege unfortunately abandons this strategy in *Grundgesetze*. For an analysis, see my [10].

[6] Cf. Evans, op. cit., who explains fundamental Ideas in terms of 'fundamental grounds of difference.'

[7] In a more complete treatment, each example would of course have to be developed in greater detail and defended against objections. My present goal is merely to sketch some promising examples in order to illustrate how the Fregean framework functions.

[8] I elaborate on this view and defend it against some natural objections in my [13].

(b) *Directions*. A direction is most directly presented by means of a line (or some other directed object) that has the direction in question. Two lines determine the same direction just in case they are parallel.
(c) *Shapes*. This case is analogous to that of directions: Shapes are most directly presented by things or figures that have the shape in question. Two such things or figures determine the same shape just in case they are congruent.
(d) *Syntactic types*. Syntactic types are most directly presented by means of their tokens. Two tokens determine the same type just in case they count (according to the relevant standards) as instantiating the same type.
(e) *Natural numbers*. A natural number is most directly presented by means of some member of a sequence of numerals. Two numerals determine the same number just in case they occupy analogous positions in their respective sequences.[9]

These examples suggest that canonical cases of singular reference are always based on two elements. First, there is an intermediary entity in terms of which the referent is most immediately presented. Let's call this the *presentation*. Second, there is a relation which specifies the condition under which two presentations determine the same referent. Let's refer to this as the *unity relation*. Finally, let's call an ordered pair $\langle u, \approx \rangle$ consisting of a presentation u and a unity relation $\approx$ applicable to this presentation a *referential attempt*. Frege's proposal is then that canonical reference is based on referential attempts. Does this proposal yield an adequate explanation of what someone's understanding of singular reference consists in?

A *formal* adequacy condition is obviously that the account be non-circular. It is easily seen that the form of our proposal allows it to be non-circular. Consider for instance the case of directions. What is proposed is that someone's understanding of an identity statement concerning directions can be explained in terms of his being suitably related to lines (in terms of which directions are presented) and having a suitable grasp of parallelism (which is the unity relation). In this case there is no threat of circularity, as we can explain what it is for someone to be suitably related to lines and to have a suitable grasp of parallelism without presupposing any prior ability to explain reference to directions. Next, we observe that there is nothing in this example that is peculiar to the case of directions. Our proposal is to explain someone's understanding of an identity statement in terms of his being suitably related to the relevant presentations and having a suitable grasp of the relevant unity relation. This explanation will of course have to include an account of what it is for a person to be suitably related to these presentations and to have a suitable grasp of this unity relation. But there is no general reason why this account should presup-

[9] I elaborate on this view in Section 3 below; related ideas are found in [15]. This view of the natural numbers as finite *ordinals* contrasts with the logicist view that the natural numbers are finite *cardinals*, individuated by Hume's Principle (which says that two numbers are identical just in case the concepts whose numbers they are equinumerous). However, both views are compatible with the Fregean account of reference. It is thus largely an empirical question which view best describes human thought about the natural numbers. See Section 3.

pose what we are trying to explain, namely reference to the sort of objects that are determined by these presentations and this unity relation.[10]

The *material* adequacy condition is that the account should capture what someone's capacity for singular reference consists in. My argument that this adequacy condition is satisfied is based on two claims. First I claim that my account explains what the subject's understanding of identity statements involving the referent consist in. Consider a representation **a** purporting to make singular reference to some object. According to my account, this representation is associated with some referential attempt $\langle u, \approx\rangle$, which specifies how the referent is presented and when two such presentations determine the same referent. By operating with this referential attempt, the subject will be able to understand any thought of the form $\ulcorner \mathbf{a} = \mathbf{b} \urcorner$, where **b** is any other representation purporting to make singular reference to an object of the kind in question. For according to my account, **b** too must be associated with some referential attempt, say $\langle v, \approx\rangle$. Moreover, we are assuming that the subject operates correctly with these representations, namely in accordance with the following principle for the identity of their semantic values:

$$[\![\mathbf{a}]\!] = [\![\mathbf{b}]\!] \leftrightarrow u \approx v \tag{SV}$$

This means that the subject has an ability to track the referent of **a** and to distinguish it from other objects of the same sort.

Next I claim that this competence is naturally described as knowing what object the representation **a** refers to. Consider for instance the case of physical bodies. Assume someone is digging in the garden, hits upon something hard with her shovel, and as a result forms the thought: This body is large. Later she hits upon something hard again, one meter away from the first encounter, and as a result forms the thought: This body is identical to that body. Finally, our subject appreciates that this identity statement is true just in case the two chunks of solid stuff that she has hit upon are spatiotemporally connected in the suitable way. It is extremely plausible to describe this capacity as a capacity to refer to physical bodies. For instance, if a robot was equipped with perception-like mechanisms and programmed so as to operate with the appropriate unity relation, it would make sense to ascribe to the robot a basic capacity for referring to physical bodies.

More generally, the unity relation $\approx$ implicitly defines a (partial) function f that maps a presentation u to the referent, if any, that u picks out. This is encapsulated in what I will call *principles of individuation*:[11]

---

[10] This is of course not to say that there cannot be *particular* cases where such an illicit presupposition exists. In fact, in Section 2.2 I will suggest that some problems encountered by Frege's Proposal are caused by the use of presentations and unity relations an adequate grasp of which *would* presuppose an ability to refer to the entities in question, thus making the account viciously circular.

[11] Principles of the form (PI) are sometimes called *two-level criteria of identity*. For two approaches based on such principles, see [17] and [20], Chapter 9.

$$f(u) = f(v) \leftrightarrow u \approx v \tag{PI}$$

Of course, when formulating principles of individuation, we philosophers make use of our own ability to refer to objects of the kind in question. But this is perfectly permissible. We are allowed to presuppose *that* we can refer to objects of the kind in question. What we are not allowed to presuppose is an explanation of what this ability *consists in*. But no such presupposition is made.

Principles of individuation are obviously closely related to what are known as *abstraction principles*, that is, to principles of the form

$$\Sigma(\alpha) = \Sigma(\beta) \leftrightarrow \alpha \sim \beta \tag{$*$}$$

where Σ is a term-forming operator, α and β are (either first- or second-order) variables, and where $\sim$ denotes an equivalence relation on the entities that α and β range over. However, the notion of an abstraction principle is a purely logical one, which applies to any principle of the form just described. By contrast, the notion of a principle of individuation is a philosophical one, which applies to a principle of the requisite form just in case it can serve in a Fregean account of reference.

This principal difference brings with it some other differences as well. For instance, the operator Σ in an abstraction principle is required to be defined on all values of the variables α and β, whereas the function f in a principle of individuation is allowed to be partial. We allow this because some presentation fail to determine referents; there are for instance spatiotemporal chunks that fail to determine unique bodies. This has consequences for the unity relation $\approx$ as well. Recall that the unity relation gives the condition under which two presentations determine the same referent. This means that any unity relation has to be symmetric and transitive. However, the unity relation will be reflexive only on those presentations that succeed in determining referents. For a referential attempt $\langle u, \approx \rangle$ succeeds in determining a referent just in case the presentation u bears the unity relation $\approx$ to itself. It follows that the domain of the partial function f is identical to the field of the unity relation $\approx$.[12] (In Section 2.2 we will encounter another difference as well, namely that the Non-Circularity Constraint puts serious restrictions on what abstraction principles can serve as principles of individuation.)

2 Challenges to the Fregean Account of Reference

In the previous section I described Frege's Proposal and outlined an attempt to carry out this Proposal. Although I argued that this attempt looks quite promising, a large number of challenging questions remain. I will now discuss four such questions,

[12] Recall that the *field* of a relation R is the set of objects which R relates. Thus, when R is dyadic, its field is the set $\{x \mid \exists y(Rxy \vee Ryx)\}$. Note that it is a theorem of first-order logic that a symmetric and transitive relation is reflexive on all objects in its field.

which are all concerned with very general features of Frege's Proposal. These are hard questions, each deserving a paper-length response of its own. All I can do here is to outline some responses and thus make it plausible that my account has the resources needed to address these worries.

2.1 Basic Reference

When a subject makes a singular reference based on a referential attempt $\langle u, \approx \rangle$, what is the nature of his relation to the presentation u? Here are three possible answers.

The first answer is that the subject's relation to the presentation u is one of full-fledged singular reference. My account of reference can then be applied again, yielding the conclusion that reference to u has to be based on some further presentation and an appropriate unity relation. Such iterated applications of my account of reference are clearly possible. For instance, in example (e) of the previous section I claimed that natural numbers are picked out by means of numeral types, which in turn are picked out by means of numeral tokens. However, on pain of a vicious regress, this cannot be a complete answer; in particular, it cannot explain how reference comes about in the first place. So at least in some cases, the subject's relation to a presentation u must be something less than full-fledged singular reference. I will say that a case of full-fledged singular reference is *basic* when it doesn't depend on any other cases of such reference. What I have argued is thus that the first answer is fine as far as it goes but that it cannot account for basic reference.

The second answer holds (like the first) that the subject's relation to the presentation u is a referential one, but insists (unlike the first) that the presentation u need not be fully individuated. Since the Fregean account of reference applies only to objects that are fully individuated, this allows us to avoid a vicious regress. For when a presentation u isn't fully individuated, the Fregean account cannot be applied to u. Instead the presentation u will contribute to a instance of basic reference.

However, it is far from clear how the not-fully-individuated presentations are to be understood. Let's focus on the clearest example of basic reference, namely reference to physical bodies. We would like to know how the required talk of spatiotemporal parts and spatiotemporal continuity can be understood in a pre-individuative manner. Probably the most developed reponse is due to Michael Dummett, who uses as his paradigm what P.F. Strawson calls 'feature-placing' sentences, such as 'It is wet here' or 'It is hot there'.[13] Such sentences involve no singular reference at all: The pronoun 'it' is merely a formal subject, not a semantic one. The same goes, according to Dummett, for the claim that 'this is continuous with that' (accompanied by two pointing gestures). This claim doesn't involve any form of singular reference to determinate parts but only feature-placing. However, it is not clear that this response delivers what we need. To begin with, this would mean that the variables u

[13] See [3], p. 217 and [4], p. 162.

and v in the principles of individuation cannot always be ordinary first-order variables; for such variables range over ordinary (and thus 'fully individuated') objects. But more seriously, it is unclear how we are to understand talk about and quantification over some class of presentations before it has been determined what it is for two such presentations to be identical or distinct.

The third answer to our question is that a subject's relation to a presentation isn't a referential one at all but something more primitive. I believe this answer provides the most promising account of basic reference. By denying that the relation between a subject and a presentation need be a referential one, this answer allows us to avoid a vicious regress. And it does so in a way that avoids any vague talk about not-fully-individuated presentations. But if the relation between a subject and a presentation isn't a referential one, what then is it? In the case of reference to physical bodies, the subject's relation to the spatiotemporal parts that serve as presentation can plausibly be taken to be a purely causal one: It is the relation that holds between the subject's perceptual system and the sum-totals of particle-instants with which the subject causally interacts in the appropriate perceptual way. In cases of reference that rely on concepts as presentations, the relation between subject and presentation is that of concept possession. Granted, it is far from clear how concept possession is to be explained; and I have no account on offer here. Even so, it is reasonably clear that no plausible account of concept possession will identify the *possession* of a concept with *reference* to that concept. This provides another example of how the relation between a subject and a presentation need not be a referential one.

2.2 The Bad Company Problem

Another problem faced by my Fregean account of reference is the need for a principled demarcation of the principles of individuation that are acceptable from those that are not. The most serious aspect of this problem has to do with *the consistency* of the underlying abstraction principles. For although the abstraction principles considered in Section 1 are consistent, it is well known that many others are not. The most famous example of this is Frege's Basic Law V, which says that the extensions of two concepts are identical just in case the two concepts are coextensive:

$$\hat{x}.Fx = \hat{x}.Gx \leftrightarrow \forall x(Fx \leftrightarrow Gx) \tag{V}$$

For in second-order logic, (V) allows us to derive Russell's paradox.[14] The problem of giving an informative characterization of the acceptable abstraction principles is known as *the Bad Company Problem.*

I am hopeful that this problem can be solved by carefully heeding the requirement that our Fregean account of reference be non-circular. Recall that this account

[14] The derivation of Russell's paradox requires Σ^1_1-comprehension. Reference [19] proves that the second-order theory with Δ^1_1-comprehension and (V) as its sole non-logical axiom is consistent.

seeks to translate the problem of explaining how reference comes about to the related but different problem of explaining what someone's understanding of identity statements concerning such objects consists in. If this translation is to constitute progress, the desired explanation of our capacity for understanding identity statements must obviously not presuppose any prior ability to explain our capacity for making singular reference to objects of the kind in question. I will refer to this as *the Non-Circularity Constraint*. Note that this constraint does not prevent us from presupposing *that* we are capable of making the relevant sort of singular reference. All it disallows is presupposing an ability to explain *what this capacity consists in*.[15]

The Non-Circularity Constraint requires that, when someone makes a referential attempt $\langle u, \approx \rangle$, it be possible for them to stand in an appropriate relation to the presentation u and have an appropriate grasp of the unity relation $\approx$ without already being capable of referring to the kind of objects reference to which we are attempting to explain. This means that the presentation u and the unity relation $\approx$ cannot involve or presuppose such objects as figure on the left-hand side of the principle of individuation.

Sets provide a nice example of how the Non-Circularity Constraint works. I believe a set is presented by means of the plurality of its elements, and that two pluralities determine the same set just in case they encompass the same objects. (The pluralities in question can be represented by means of plural variables, as suggested by George Boolos.[16]) We may also need to add that only pluralities that satisfy some condition of 'limitation of size' succeed in determining sets. Let's write $\textsc{Form}(uu, x)$ for the claim that the objects uu form a set x; that is, that there is a set x whose elements are precisely the objects uu. My claim is then that sets are individuated as follows

$$\textsc{Form}(uu, x) \wedge \textsc{Form}(vv, y) \rightarrow (x = y \leftrightarrow uu \equiv vv) \qquad \text{(Id-Sets)}$$

where '$uu \equiv vv$' is a formalization of the claim that the pluralities uu and vv encompass precisely the same objects.[17] Let's now apply the Non-Circularity Constraint. This constraint requires that we be able to refer to the individual objects that make up a plurality *before* we can refer to the set that this plurality forms. This means that a set cannot contain itself as an element. More generally, the constraint can be seen to give rise to the set-theoretic axiom of Foundation.

Extensions, on the other hand, are individuated by Frege's Basic Law V: They are presented by means of concepts and are identical just in case the two presenting concepts are coextensive. But unless (V) is restricted in some way it will lead straight to paradox, as Russell discovered. My hypothesis is that the problem with the unrestricted version of (V) is that violates the Non-Circularity Constraint: that its

---

[15] Recall that this distinction played an important role in my defense of (PI) towards the end of Section 1.

[16] See [1]; for an introduction, see [11].

[17] For those familiar with plural logic, this claim can be formalized as $\forall z(z \prec uu \leftrightarrow z \prec vv)$.

right-hand side somehow presupposes an understanding of what we are attempting to explain on the left-hand side, namely what it is for two extensions to be identical.

One way of developing this suggestion is as follows. Say that an abstraction principle is *predicative* when its unity relation doesn't quantify over the kinds of entities to which its left-hand side purports to refer, and that it is *impredicative* otherwise. For instance, the abstraction principle that specifies how directions are individuated from lines,

$$d(l_1) = d(l_2) \leftrightarrow l_1 \parallel l_2 \tag{D}$$

is predicative, as there is no reference to, or quantification over, directions on its right-hand side. By contrast, (HP) and (V) are both impredicative, as their right-hand sides quantify over precisely the kinds of object that their left-hand sides attempt to individuate. Could it be that the unity relations involved in impredicative abstraction principles are illegitimate because they presuppose an ability to individuate the objects in question, thus violating the Non-Circularity Constraint?

The following argument appears to suggest that this is indeed so. Let $(*)$ be an impredicative abstraction principle. The impredicativity means that the right-hand side of $(*)$ quantifies over the very sort of objects, say Fs, reference to which we are attempting to explain. Now it appears that in order to understand a quantified formula, one must be able to understand an arbitrary instance of this formula. If this is right, it follows that one cannot understand the right-hand side of $(*)$ unless one is already able to understand instances, with respect to Fs, of the formulas to which its quantifiers attach. But in order to understand such instances, one needs to be capable of making singular reference to Fs. This means that use of an impredicative abstraction principle as a principle of individuation is circular: it presupposes precisely what it attempts to explain, namely an understanding of singular reference to the relevant sort of objects.[18]

So if this argument were sound, only predicative abstraction principles would be permitted as principles of individuation. However, a crucial step in this argument fails: It is *not* the case that, in order to understand a quantified formula, one needs to be capable of understanding an arbitrary instance. We *can* understand universal generalizations without being able to understand an arbitrary instance. For instance, we can understand (and even appreciate the truth of) the claim that all objects are self-identical, although there are no doubt many kinds of object to which we will never be capable of referring. What does the work in such cases is that we have an absolutely general understanding of what it is for the relevant condition to hold of an object. This is what enables us to tell that *any object whatsoever* is self-identical, including objects of which we will never gain more substantive knowledge.

We therefore need a better analysis of what a quantified formula presupposes. I will here focus on the presuppositions carried by the open formulas (or *conditions* as

[18] A similar argument is developed in [6], Section II.5, whereas the opposite view is defended in [22] and [23].

I will also call them) that define concepts and extensions. In particular, what do such conditions presuppose about the identities of the objects on which they are defined? Since what a condition does is distinguish between objects—those of which it holds and those of which it doesn't—this notion of presupposition should be spelled out in terms of what distinctions the condition makes. Now, a condition can only presuppose an object if it is able to distinguish this object from other known objects. I therefore propose that we analyze what it is for a condition to presuppose only entities that are already individuated in terms of the condition's not distinguishing between entities not yet individuated. This proposal can be made mathematically precise as follows. Consider permutations π that fix all objects already individuated and that respect all relations already individuated in the sense that for each such relation R we have

$$\forall x_0 \cdots \forall x_n (Rx_0 \cdots x_n \rightarrow R\pi x_0 \cdots \pi x_n).$$

My proposal is then that a condition $\phi(u)$ presupposes only entities that are already individuated just in case $\phi(u)$ is invariant under all such permutations. The 'Russell condition' $\exists F(u = \hat{x} \cdot Fx \wedge \neg Fu)$ provides an interesting example. This condition is easily seen to violate the present analysis of the Non-Circularity Constraint; for this condition is sensitive to the characteristics of the object assigned to the variable u. So this condition cannot be allowed to define a concept or an extension.

When a condition does satisfy the Non-Circularity Constraint, there is no obvious philosophical reason why it should not define an extension. For this extension has been individuated in a non-circular way. Nor is there any mathematical reason why such a formula should not define a extension.[19] To see this, begin by observing that such a condition defines a *concept*. Next I claim that any concept individuated in accordance with the Non-Circularity Constraint can be assigned an extension. For each such concept can be represented by means of one of the objects not yet individuated. Since these concepts don't distinguish between objects not yet individuated, it doesn't matter which representative we choose. So if we assume that there are as many objects not yet individuated as there are sets (an assumption that can easily be modeled within set theory), there will be enough objects to represent all the concepts definable by conditions expressible in any reasonable language. And we can carry out this process as many times as there are sets. I therefore conclude that when the individuation of concepts is subjected to the Non-Circularity Constraint, the resulting concepts do indeed have extensions in accordance with Basic Law V.

2.3 Compositionality

I now turn to a worry about the compatibility of my Fregean account of reference with the principle of compositionality. This worry is based on the following ob-

[19] What follows is an intuitive presentation of ideas developed in more detail in Sections 7 and 8 of my [14].

servation. According to the principle of compositionality, the meaning of anything complex is to be explained in terms of the meanings of its constituent parts. So here the order of explanation goes from what is complex to what is simpler. But according to my account of reference, the referentiality of a singular representation is partially explained in terms of the meaningfulness of identities involving this representation. So here the order of explanation goes in the opposite direction: from what is absolutely simple to what is more complex. Given these opposite orders of explanation, one may think my account conflicts with the principle of compositionality.

One way *not* to resolve this apparent conflict would be by claiming that, whereas my account is only concerned with thought, the principle of compositionality governs only linguistic meaning. This response is unacceptable because an analogous principle of compositionality applies to the contents of thoughts.

However, there is another respect in which the principle of compositionality and my Fregean account of reference do have completely different concerns. To see this, we need to distinguish between what I will call *semantics* and *meta-semantics*.[20] Semantics standardly takes the form of a theory of *semantic values*, where the semantic value $[\![\mathbf{E}]\!]$ of an expression $\mathbf{E}$ is the contribution that this expression makes to the truth-values of sentences in which it occurs.[21] Following Frege, the semantic value of a sentence is often taken to be just its truth-value, and the semantic value of a proper name, its referent. If so, it follows that the semantic value of a one-place predicate must be a function from objects to truth-values. Following Frege again, it is argued that semantic values are subject to a *principle of compositionality*, according to which the semantic value of a complex expression is determined as a function of the semantic values of its individual sub-expressions. For instance, the semantic value of an atomic sentence $\mathbf{P}(\mathbf{a_1}, \ldots, \mathbf{a_n})$ is functionally determined as $[\![\mathbf{P}]\!]([\![\mathbf{a_1}]\!], \ldots, [\![\mathbf{a_n}]\!])$.

Meta-semantics, on the other hand, is concerned with what is involved in an expression's having the various semantic properties that it happens to have, such as its semantic structure and its semantic value. The relation between semantics and meta-semantics can be compared with that between economics and what we may call *meta-economics*. Economics is concerned with the laws governing money; for instance, that an excessive supply of money leads to inflation. Meta-economics, on the other hand, is concerned with what is involved in various objects' having monetary value; for instance, what makes it the case that a piece of printed paper can be worth 100 euros. Clearly, such facts cannot be primitive but must have an explanation.

The meta-semantic questions that we are currently interested in concern singular terms and representations. How does the relation of reference come about? What makes it the case that a 'dead' syntactic object—some ink marks on paper or neural configurations in the brain—'reach out' to some referent with which the singular

---

[20] My distinction between semantics and meta-semantics is thus the same as Stalnaker's distinction between 'descriptive' and 'foundational' semantics. See e.g. [18].
[21] I will use boldface for all meta-linguistic variables.

term bears no intrinsic connection? Just like the fact about the value of a banknote, this semantic fact cannot be primitive but must have an explanation.

I would instead like to suggest that our distinction between semantics and meta-semantics provides the key to resolving the apparent conflict between my Fregean account of reference and the principle of compositionality. The principle of compositionality is concerned with the assignment of semantic values to complex expressions and thus belongs to semantics. My Fregean account of reference, on the other hand, is concerned with what is involved in an expression's having the various semantic properties it happens to have and thus belongs to meta-semantics. Since the principle of compositionality and our Fregean account of reference have completely different concerns, there is no conflict.

However, the most popular response among philosophers who seek to use Frege's Context Principle to explain reference has been to concede that the apparent conflict is genuine and therefore to argue that the principle of compositionality has to be rejected or at least weakened.[22] But rejecting or weakening the principle of compositionality is obviously a steep price to pay. Why, then, have so many philosophers found this response inevitable? I believe the answer has to do with a dangerous ambiguity in the wording of Frege's proposal. 'Since it is only in the context of a sentence that words have any meaning,' Frege writes, 'our problem becomes this: *To define the sense of a sentence in which a number word occurs*.'[23] This is ambiguous between a semantic and a meta-semantic reading. On the semantic reading, our task is to *specify* the meaning or sense of identity statements in which number words occur. But on the meta-semantic reading, our task is to *explain what makes it the case that such identity statements have the meanings that they happen to have*.

Frege's proposal has traditionally been interpreted along the lines of the semantic reading. It is then natural to assume that what Frege proposes is that the meaning of problematic identity statements be given by a 'reductive' truth-condition

$$\ulcorner \mathbf{a} = \mathbf{b} \urcorner \text{ is true iff } u \approx v \qquad \text{(T-Red)}$$

where **a** and **b** are representations associated with referential attempts $\langle u, \approx \rangle$ and $\langle v, \approx \rangle$ respectively. On this reading there will indeed be a conflict with the principle of compositionality. For according to this principle, the semantic value of an atomic sentence $\mathbf{P}(\mathbf{a_1}, \ldots, \mathbf{a_n})$ is functionally determined as $[\![\mathbf{P}]\!]([\![\mathbf{a_1}]\!], \ldots, [\![\mathbf{a_n}]\!])$. Applied to the identity $\ulcorner \mathbf{a} = \mathbf{b} \urcorner$, this yields something different from (T-Red), namely the completely trivial truth-condition

$$\ulcorner \mathbf{a} = \mathbf{b} \urcorner \text{ is true iff } [\![\mathbf{a}]\!] = [\![\mathbf{b}]\!]. \qquad \text{(T-Triv)}$$

And any further or alternative semantic analysis is out of the question, given that the terms **a** and **b** are supposed to be semantically simple. Moreover, the truth-condition

---

[22] See [8] and [21].

[23] *Grundlagen* Section 62; my italics.

(T-Triv) will be of absolutely no use in the project of explaining some problematic form of reference. For the right-hand side of (T-Triv)—unlike that of (T-Red)—involves precisely the sort of reference that we are attempting to explain.

Faced with this choice between the reductive truth-condition (T-Red), which allows the explanatory project to progress, and the trivial one (T-Triv), on which the explanation cannot even get started, it is of course tempting to insist that it is the former that gives the meaning of the identity statement, and that if this conflicts with the compositionality of meaning, then so much the worse for this principle of compositionality. This appears to have been Frege's view in *Grundlagen*, where he talks about the right-hand side being a 'recarving' of the meaning of the left-hand side.[24] This 'recarving thesis' is explicitly endorsed by prominent contemporary defenders of Fregean ideas about reference, such as Bob Hale and Crispin Wright.[25]

However, I have insisted throughout this paper that Frege's proposal is of meta-semantic nature (although it is in only in this sub-section that I have explicitly labeled it as such). I am therefore under no pressure to say that (T-Red) gives the meaning of an identity statement. I can instead maintain that the only semantically generated truth-condition for an identity statement is the trivial one (T-Triv). What Frege proposes is rather an account of what a subject's understanding of an identity statement consists in. And as we have seen, this account involves the principle for the identity of their semantic values:

$$[\![\mathbf{a}]\!] = [\![\mathbf{b}]\!] \leftrightarrow u \approx v \tag{SV}$$

When this principle (SV) is combined with the trivial truth-condition (T-Triv), we do indeed get the reductive one (T-Red), which now emerges as a hybrid of semantic and meta-semantic facts.

A worry related to the one about compositionality concerns the compatibility of my Fregean account of reference with the semantic thesis that names and their mental counterparts are rigid designators. Consider a representation **a** associated with a referential attempt $\langle u, \approx \rangle$. Let f be the function determined from $\approx$ in accordance with the Principle of Individuation (PI). I have argued that **a** refers, if at all, to the object $f(u)$. One may then wonder whether my view isn't committed to some version of the now widely rejected *descriptivist* view of names, which identifies the semantic value of a name with that of some description associated with the name. Specifically, am I not committed to identifying the meaning of **a** with that of the description 'the f of u,' with the result that **a** isn't a rigid designator after all? (For instance, a chunk of physical matter may be part of different bodies in different possible worlds.)

The above discussion provides the resources needed to dismiss this worry. On my proposal, the nature of the function-argument structure $f(u)$ is entirely meta-semantic. The semantic value of **a**, if any, will be *the object* $f(u)$. How this referent

---

[24] See [7], Section 64.

[25] See [8] and [21].

is determined is entirely a meta-semantic matter, and thus of no immediate semantic significance. As far as semantics is concerned, **a** is a simple term or representation whose semantic value is just an object.

In fact, this view of the nature of the function-argument structure $f(u)$ enjoys independent evidence. Semantic structure is by and large accessible to consciousness; otherwise we wouldn't know or be rationally responsible for what we say and think. But someone can understand reference to shapes and bodies without having any conscious knowledge of how such reference is structured. Someone's competence with this structure may be located wholly at a subpersonal level. This is evidence that the structure isn't semantic. And if that is right—that is, if the function-argument structure of my account of reference isn't of semantic nature—then my account will be fully compatible with the rigidity thesis and in no danger of collapsing back into descriptivism (which is a semantic thesis).

2.4 Existence and Uniqueness of Referents

I now turn to two final worries about principles of individuation. Let I be a class of presentations and $\approx$ a unity relation defined on I. Assume we stand in an adequate relation to both the presentations and the unity relation. Then I am committed to the claim that I and $\approx$ implicitly define a partial function f such that

$$f(u) = f(v) \leftrightarrow u \approx v \tag{PI}$$

and such that f is defined on all presentations in the field of the relation $\approx$.

Two worries arise concerning this view. Firstly, how do we know that there exist objects of the sort that are supposed to be in the range of the partial function f? For instance, assuming that lines exist and that each is parallel to itself, how do we know that their directions exist? To answer such skeptical questions, we need to analyze claims to the effect that Fs exist, where F is the concept implicitly defined by the class of presentations I and the unity relation $\approx$. These claims come out true provided there are semantic values falling under the concept F. And a semantic value falls under the concept F just in case it can be presented by a referential attempt $\langle u, \approx \rangle$ which is successful in the sense that $u \approx u$. But this is just what we have assumed. So the desired kind of semantic values *will* exist, and the claim that Fs exist will therefore be true.

It may be objected that this response trivializes ontological questions. How can what I have just described be all that is required for an object to exist? I admit that some of the objects I have talked about are very 'light-weight,' in the sense that their existence doesn't amount to very much. (I will have more to say about this in the next section, where I discuss mathematical objects.) But other objects are less 'light-weight.' For instance, the condition for a referential attempt to pick out a physical body is what we would expect: the presentations u must be a chunk of solid stuff, and this chunk must be related to other chunks in such a way as to define a unified whole of solid stuff that moves as a unit.

It may also be objected that my response to the first worry opens the door to all kinds of strange and unusual objects which we normally never talk about. For instance, let the class of presentations consist of people and the unity relation be siblinghood. On my account this is a coherent form of reference. Does this show that my account is committed to an extravagant and implausible ontology? Although I admit that my account is committed to a generous ontology, I believe it also has the resources needed to explain why this ontology is harmless. In particular, the apparent implausibility of the unusual objects which my account countenances can be adequately explained by the fact that canonical reference to such objects is very different from any kind of reference in which we ordinarily engage.

The second worry is that there may be *more than one* equally good candidates for playing the role of the function f. After all, whenever a function f satisfies (PI), then so does $\sigma \circ f$, where σ is any permutation of the Fs. Can we point to some feature of the intended function f that distinguishes it from its unintended rivals? Unfortunately, a more explicit characterization of the function f is out of the question. For a more explicit characterization—whether in the form of an algorithm or simply as a set of ordered pairs—would presuppose that the objects in the range of this function can be referred to in some way *other* than by means of this function. But on my account, there is no such alternative way of referring to these objects, since the function f plays an essential role in canonical reference to objects of the sort in question. So the function f cannot be given a more explicit characterization than that given by (PI).

Fortunately, this very fact can also be used to assuage the worry. Assume someone successfully uses a referential attempt $\langle u, \approx \rangle$. Assume some philosopher studies this referential attempt and asks which of two candidate referents this referential attempt has singled out. How can this question be answered? If my theory is completely general, the philosopher must herself pick out the two candidate referents by means of referential attempts. Since two referential attempts can pick out the same referent only if their unity relations are compatible, we may assume these referential attempts to be of the form $\langle v_i, \approx \rangle$. But then our philosopher can determine which candidate is right by determining whether $u \approx v_i$ for either i.

3 Referring to the Natural Numbers

I will now apply some of the ideas of the previous sections in an attempt to explain how we manage to refer to the natural numbers (where by 'we' I will mean ordinary speakers of English with at least basic competence in arithmetic). Since this is an account of an ability that we have, it will be relevant to consider some psychological facts about us. My account is based on two claims about our ordinary thought and talk about natural numbers. I begin by explaining and defending these two claims.

My first claim is that we regard the natural numbers as finite *ordinals*, individuated by their position in a well-ordering, rather than as finite *cardinals*, individuated by the cardinality of the sets or concepts whose numbers they are.[26] I grant that *one* way of thinking and talking about the natural numbers is by means of expressions of the form 'the number of Fs', as the cardinal-based approach will have it. What I deny is that this is our most fundamental or direct way of thinking and talking about the natural numbers. I offer four arguments for this claim.

1 *A phenomenological argument.* If there are seven apples on a table, we can think of the number seven as the number of apples on the table. Or (following Frege) we can think of seven as the number of numbers less than or equal to six. But neither or these ways of thinking of the number seven *feels* very direct or explicit. A more direct and explicit way of thinking of seven is by means of the standard numeral '7' or by means of any other numeral occupying the seventh position in a system of numerals with which we are familiar.
2 *An argument from the philosophy of language*. Expressions of the form 'the number of Fs' aren't singular terms but definite descriptions. It is for instance easily seen that such expressions aren't rigid (although the ones couched entirely in mathematical language will be *de facto* rigid, in the sense that the description is true of the same object in all possible worlds). But it is doubtful that there can be a kind of objects to which reference is possible *only* by means of definite descriptions. In order to refer to an object by means of a definite description, there must be some other, more direct way of thinking of the object.
3 *An argument from the number zero*. If our most fundamental way of thinking about the natural numbers is as cardinals, then zero would have been the most obvious and immediate number of them all. If on the other hand our most fundamental way of thinking about the natural numbers is as ordinals, then zero would be no more obvious or immediate than the negative numbers. This latter hypothesis accords much better with the late stage at which zero was admitted into mathematics as a number in good standing.
4 *A technical argument*. In [12] I showed that on the cardinal-theoretic approach, the standard proofs of some basic predicative arithmetical truths require impredicative second-order comprehension. This is not only unnatural but makes these very elementary arithmetical truths depend on much too theoretical principles. This problem goes away when the natural numbers are regarded as ordinals.

My second claim is that it is part of ordinary arithmetical competence that the natural numbers are *notation independent*, in the sense that they can be denoted by different systems of numerals. Indeed, even people with very rudimentary knowledge of arithmetic know that the natural numbers can be denoted by ordinary decimal numerals, by their counterparts in written and spoken English (and in other natural languages), and by sequences of strokes (perhaps grouped in fives). Many people also know alternative systems of numerals such as the Roman numerals and the nu-

---

[26] The view that the natural numbers are finite cardinals is defended by the classical logicists Frege and Russell, as well as by contemporary neo-logicists. Although this view is compatible with my Fregean account of reference, I don't think it gives an accurate description of our ordinary arithmetical thought and talk.

merals of position systems with bases other than ten, such as binary and hexadecimal numerals. To accommodate such alternative systems of numerals, we need a *general* condition for two numerals to denote the same number and for one numeral to be related to another in such a way that the number denoted by the former immediately precedes the number denoted by the latter.

I will here take a numeral to be any object that occupies a position in a well-ordering. In fact, since it is convenient to make the well-ordering explicit, I will take a numeral to be an ordered pair $\langle u, R\rangle$, where u is the numeral proper and R is the well-ordering in which u occupies a position. On this very liberal view of the matter, the numeral proper u need not be a syntactical object, at least not in any traditional sense. For instance, if a pre-historic shepherd counts his sheep by matching them with cuts in a stick, then these cuts count as numerals. Moreover, since R can be any well-ordering, these numerals refer to ordinal numbers but not necessarily to finite ones.

Our first task is to describe the equivalence relation that holds between two such pairs $\langle u, R\rangle$ and $\langle u', R'\rangle$ when they determine the same number. This equivalence relation must obviously be a matter of the two objects u and u' occupying analogous positions in their respective orderings, for instance, that both occupy the 17th position. More formally, $\langle u, R\rangle$ and $\langle u', R'\rangle$ are equivalent just in case there exists a relation C which

- is an order-preserving correlation of initial segments of R and R'
- is extensive enough to have both u and u' in its field
- is such that $C(u, u')$.[27]

Let $\langle u, R\rangle \approx \langle u', R'\rangle$ symbolize that the two ordered pairs are equivalent in this sense. Ordinal numbers are then individuated by the following abstraction principle:

$$N\langle u, R\rangle = N\langle u', R'\rangle \leftrightarrow \langle u, R\rangle \approx \langle u', R'\rangle \qquad \text{(Id-}N\text{)}$$

Note that the numbers to which the numerals are mapped are not equivalence classes of numerals but form their own category of objects. (Likewise, physical bodies are *sui generis* objects rather than equivalence classes of chunks of physical stuff.) Let '$O(x)$' be a predicate that holds of all and only the objects that can be presented in this way.

Next we define a predecessor relation $P^{\#}$ on such pairs by letting $P^{\#}(\langle u, R\rangle, \langle u', R'\rangle)$ just in case there is an R'-predecessor v of u' such that $\langle u, R\rangle \approx \langle v, R'\rangle$. It is easily verified that $P^{\#}$ is a congruence with respect to $\approx$ (that is, that $P^{\#}$ doesn't distinguish between $\approx$-equivalent objects). This means that $P^{\#}$ induces a predecessor relation P on the ordinal numbers themselves, defined by

$$P(N\langle u, R\rangle, N\langle u', R'\rangle) \leftrightarrow P^{\#}(\langle u, R\rangle, \langle u', R'\rangle). \qquad \text{(Def-}P\text{)}$$

[27] The well-foundedness of the relations R and R' guarantees that there is an order-preserving correlation of initial segments extensive enough to include at least one of u and u'. And this is all we need to determine whether the condition in the main text is met.

Finally, following the practice of ordinary counting, we let 1 be the first number. For instance, 1 may be presented as $N\langle\text{‘1’}, D\rangle$, where D is the familiar well-ordering of decimal numerals.

These definitions enable us to establish some of the basic axioms for ordinal numbers. We begin with the easier ones.

(O1) $O(1)$

(O2) $\neg\exists x\, P(x, 1)$

(O3) $P(x, y) \wedge P(x', y) \to x = x'$

(O4) $P(x, y) \wedge P(x, y') \to y = y'$

(O5) $\forall X[\exists x(Ox \wedge Xx) \to \exists x(Ox \wedge Xx \wedge \forall y((Oy \wedge Xy \wedge x \neq y) \to P(x, y)))]$

The proof of (O1) is trivial. For (O2), let n be any ordinal number. Then there is some presentation $\langle u, R\rangle$ such that $n = N(\langle u, R\rangle)$. If $P(n, 1)$, then u would come before the first element in a well-ordering. But this is impossible. (O3) follows from the observation that any two numerals which immediately precede a third are equivalent. (O4) follows from the observation that any two numerals which immediately succeed a third are equivalent. (O5) follows from the observation that the numerals are well-ordered.

However, I have not yet said anything very substantial about how many ordinals there are. For the purpose of describing the natural numbers (which I have identified with the finite ordinals), the only such principle we need is *the Successor Axiom*:

$$\forall x(O(x) \to \exists y P(x, y)) \tag{O6}$$

The Successor Axiom would follow immediately from a corresponding principle about presentations:

$$\forall\langle u, R\rangle\, \exists\langle u', R'\rangle\ P^{\#}(\langle u, R\rangle, \langle u', R'\rangle)$$

However, it is doubtful that this principle about presentations has the required epistemological status. I therefore instead adopt the following weaker principle:

$$\Box\forall\langle u, R\rangle\Diamond\exists\langle u', R'\rangle\ P^{\#}(\langle u, R\rangle, \langle u', R'\rangle)$$

This modified principle is extremely plausible. For assume we're given a presentation $\langle u, R\rangle$. Then it is possible that there should be some object u' not among the relata of R. Let R' be the result of adding the pair $\langle u, u'\rangle$ to the initial segment of R ending with u. Then $\langle u', R'\rangle$ is as desired. Moreover, combined with a claim to the effect that ordinals exist by necessity, this modified implies the Successor Axiom.

Finally, we need to specify some condition of finitude with which to restrict the ordinals such that we get all and only the natural numbers. I claim that this condition is simply that mathematical induction be valid of the natural numbers. That is, an ordinal n is a natural number (in symbols: $\mathbb{N}n$) just in case the following open-ended schema holds:

$$\phi(1) \wedge \forall x \forall y[\phi(x) \wedge P(x, y) \rightarrow \phi(y)] \rightarrow \phi(n) \tag{MI}$$

Our characterization of the natural numbers has thus allowed us to prove all the Dedekind-Peano axioms.

4 The Metaphysical Status of the Natural Numbers

Does the account of reference to the natural numbers that I have outlined tell us anything about their metaphysical status? In particular, does the account favor either a platonist or a nominalist interpretation of the language of arithmetic?

Before addressing the question of platonism directly, it is useful to explain a fundamental difference between physical bodies and natural numbers having to do with the ways in which they possess properties. Consider the question whether a physical body x has some property, say being round. To answer this question, it's not sufficient to consider any proper part of x. Whether a body is round isn't determined by any of its proper parts but information is needed about the entire body. And there is nothing unusual about this case. It is in general true that, in order to determine whether a body x has some property G, one needs information about *many* parts of x. The question whether a body has some property G cannot in general be reduced to a question about any *one* of its proper parts. This means that a body can have properties in an irreducible way, that is, in a way that isn't reflected in any properties of any one of its proper parts.

The situation is very different with natural numbers. Consider the question whether a natural number n has some mathematical property G, say the property of being even. In this case a standard presentation of n by some numeral (say a standard decimal numeral **n**) suffices to answer the question. The question whether the natural number n possesses the property G can be reduced to a question about the numeral **n** by which n is presented. Here's why. All the usual arithmetical properties are definable (in second-order logic) from the predecessor relation P. And as (Def-P) shows, the question whether P holds between two natural numbers is itself reducible to the question whether the relation $P^\#$ holds between numerals. Natural numbers are therefore 'impoverished' compared to numerals. For whenever a natural number n possesses some property, its doing so is inherited from the fact that the numerals that present n possess some associated property. Natural numbers are therefore 'thinner' than the numerals that present them. This opens for a form of reductionism about natural numbers: Questions about such objects can be reduced to questions about their presentations.[28]

[28] For reasons that I cannot here go into, I think that the same hold of other mathematical objects as well.

Given this reductionism, does it still make sense to say that numerals refer to natural numbers? I believe this question is best understood as the question whether it still make sense to ascribe semantic values to numerals. I will now argue that this does still make sense. One observation that supports this claim is the following. It is always the default assumption that a syntactically uniform class of expressions, such as the class of singular terms, should have a uniform semantic analysis. Now, when we analyze English and the language of arithmetic, singular terms such as '5' and '1001' seem to function just like terms such as 'Alice' and 'Bob'. The default assumption is therefore that all these terms function in similar ways. Since singular terms such as 'Alice' and 'Bob' clearly have semantic values (namely the physical bodies that they refer to), this provides at least some reason to think that arithmetical singular terms such as '5' and '1001' have semantic values as well.

It may be objected that this default assumption is overridden by our discovery that questions about natural numbers can be reduced to questions about the associated numerals. Since this reduction shows that it suffices to talk about the numerals themselves, the objection continues, there is no need to ascribe any sort of semantic values to numerals. However, this objection will succeed *only if the structure responsible for the reduction that we have discovered is also the kind of structure that matters for semantic analysis*. But I argued in Section 2.3 that this condition is not met, because the structure responsible for the reduction isn't semantically accessible but is entirely of meta-semantic nature.

If these arguments are on the right track, then it *does* make sense to ascribe semantic values to numerals. It is therefore perfectly true to say that natural numbers exist. So clearly my analysis isn't a nominalist one. However, my analysis is also far removed from a traditional platonist conception of mathematics. For although I have argued that numerals do have abstract semantic values, I have also argued that this fact allows of a deeper, non-semantic analysis on which any such reference to mathematical objects disappears. To ascribe to a language a certain semantic structure is to ascribe to it a certain kind of pattern of regularities in the ways truth-values of sentences are determined—a pattern of a kind that exists throughout human language and thought, and which has some important linguistic and psychological features. But this pattern may in turn rest on a deeper pattern on which reference to mathematical objects drops out. Reductionism about reference can therefore be incorrect on a semantic analysis, although correct on a different, more thoroughgoing form of analysis.

Where does this leave the question of mathematical platonism? If by 'mathematical platonism' is meant simply the view that there are true sentences some of whose semantic values are abstract, then my view is obviously a platonist one. But given how light-weight these semantic values are, this may be more of a reason to reject the above definition of mathematical platonism than for hard-line platonists to declare victory.

Acknowledgments Thanks to Matti Eklund, Anthony Everett, Agustín Rayo, Stewart Shapiro, and Timothy Williamson for discussion and comments on earlier versions of this paper. Thanks also to audiences at Uppsala University and the University of Oxford, where earlier versions of this paper were presented.

References

1. George Boolos. To Be is to Be a Value of a Variable (or to Be Some Values of Some Variables). *Journal of Philosophy*, 81(8):430–449, 1984. Reprinted in [2].
2. George Boolos. *Logic, Logic, and Logic*. Harvard University Press, Cambridge, MA, 1998.
3. Michael Dummett. *Frege: Philosophy of Language*. Harvard University Press, Cambridge, MA, second edition, 1981.
4. Michael Dummett. *Frege: Philosophy of Mathematics*. Harvard University Press, Cambridge, MA, 1991.
5. Gareth Evans. *Varieties of Reference*. Oxford University Press, Oxford, 1982.
6. Kit Fine. *The Limits of Abstraction*. Oxford University Press, Oxford, 2002.
7. Gottlob Frege. *Foundations of Arithmetic*. Blackwell, Oxford, 1953. Transl. by J.L. Austin.
8. Bob Hale. *Grundlagen* Section 64. *Proceedings of the Aristotelian Society*, 97(3):243–61, 1997. Reprinted with a postscript in [9].
9. Bob Hale and Crispin Wright. *Reason's Proper Study*. Clarendon, Oxford, 2001.
10. Øystein Linnebo. Frege's Proof of Referentiality. *Notre Dame Journal of Formal Logic*, 45(2):73–98, 2004.
11. Øystein Linnebo. Plural Quantification, 2004. In *Stanford Encyclopedia of Philosophy*, available at http://plato.stanford.edu/entries/plural-quant/.
12. Øystein Linnebo. Predicative Fragments of Frege Arithmetic. *Bulletin of Symbolic Logic*, 10(2):153–74, 2004.
13. Øystein Linnebo. To Be Is to Be an *F*. *Dialectica*, 59(2):201–222, 2005.
14. Øystein Linnebo. Sets, Properties, and Unrestricted Quantification. In Agustín Rayo and Gabriel Uzquiano, editors, *Unrestricted Quantification: New Essays*, pp. 149–178. Oxford University Press, Oxford, 2006.
15. Charles Parsons. Ontology and Mathematics. *Philosophical Review*, 80:151–176, 1971. Reprinted in [16].
16. Charles Parsons. *Mathematics in Philosophy*. Cornell University Press, Ithaca, NY, 1983.
17. John Perry. The Two Faces of Identity. In *Identity, Personal Identity, and the Self*. Hackett, Indianapolis, IN, 2002.
18. Robert Stalnaker. On Considering a Possible World as Actual. *Proceedings of the Aristotelian Society*, Suppl. vol. 65:141–156, 2001.
19. Kai Wehmeier. Consistent Fragments of *Grundgesetze* and the Existence of Non-Logical Objects. *Synthese*, 121:309–28, 1999.
20. Timothy Williamson. *Identity and Discrimination*. Blackwell, Oxford, 1990.
21. Crispin Wright. The Philosophical Significance of Frege's Theorem. In Richard Heck, editor, *Language, Thought, and Logic. Essays in Honour of Michael Dummett*. Clarendon, Oxford, 1997. Reprinted in [9].
22. Crispin Wright. Response to Michael Dummett. In Matthias Schirn, editor, *Philosophy of Mathematics Today*. Clarendon, Oxford, 1998.
23. Crispin Wright. The Harmless Impredicativity of $N^{=}$ (Hume's Principle). In Matthias Schirn, editor, *Philosophy of Mathematics Today*. Clarendon, Oxford, 1998.

The Measure of Scottish Neo-Logicism

Stewart Shapiro

The Scottish neo-logicist project began with the development of arithmetic in Wright [35]. It was bolstered by Bob Hale [15], and continues through many extensions, objections, and replies to objections (see [18]). The purpose of this paper is to assess the program along several different fronts. There are two preliminaries. The first is to briefly describe the program, along with its more or less standard motivations, and the second is to indicate the various yardsticks (or meter sticks) that I use to take the measure of the program.

1 The Development

The formal part of Scottish neo-logicism is to develop branches of established mathematics from abstraction principles in the form:

$$(\mathbf{ABS}) \quad \forall a \forall b(\Sigma(a) = \Sigma(b) \equiv E(a, b)),$$

where a and b are variables of a given type, usually either individual objects or concepts, Σ is a higher-order operator, denoting a function from items of the given type to individual objects, and E is an equivalence relation over items of the given type. Kit Fine [12] provides a wealth of detail on the mathematical possibilities and limitations of the program, so conceived, as well as considerable philosophical discussion of its merits.

Two concepts F, G, are equinumerous if there is a one-to-one correspondence between the objects falling under F and the objects falling under G. On a properly set table, for example, the plates are equinumerous with the napkins, and also with the water glasses. Gottlob Frege showed how to define equinumerosity using the resources of what is now known as second-order logic, and without explicitly presupposing the natural numbers. Equinumerosity is an equivalence relation on

S. Shapiro (✉)
Department of Philosophy, The Ohio State University, Columbus, Ohio 43210, USA;
Arché Research Centre, University of St. Andrews, St. Andrews, Scotland KY16 9AL
e-mail: shapiro.4@osu.edu

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_4,

concepts. His *Grundlagen der Arithmetik* [13, Section 63] contained a principle stating a more or less obvious connection between cardinal numbers and the relation of equinumerosity. It is in the form (ABS):

> For any concepts *F, G*, the number of *F* is identical to the number of *G* if and only if *F* and *G* are equinumerous.

Wright dubbed the principle $N^=$, but it has subsequently been called *Hume's Principle* (HP). Variations of HP underlie just about any theory of cardinality.

For reasons we need not get into here, Frege was not satisfied with HP as a foundation of arithmetic, or as a definition of (cardinal) number. Instead, he gave an explicit definition of the individual numbers, in terms of extensions: the number of *F* is defined to be the extension of the concept "equinumerous with *F*". One key principle in Frege's account of extensions is the infamous Basic Law V, paraphrased thus:

> For any concepts *F, G*, the extension of *F* is identical to the extension of *G* if and only if for every object *a, Fa* if and only if *Ga*.

This, too, is in the form (ABS). Frege's logicism came down in ruins when it was discovered that Basic Law V is inconsistent, via Russell's paradox.

Along with a number of authors, Wright pointed out that Frege's development of arithmetic [13, 14] contains the essentials of a derivation, in second-order logic, of the standard Peano-Dedekind axioms from Hume's Principle. Indeed, the only substantial use of Basic Law V in Frege's presentation is to derive Hume's Principle from the explicit definitions. The derivation of the usual axioms of arithmetic from HP is now known as *Frege's theorem*. Moreover, Hume's Principle is consistent if second-order arithmetic is. See, for example, Parsons [23], Wright [35], Hodes [20], and Boolos [3].

The Scottish neo-logicist proposes to bypass the theory of extensions and to use Hume's Principle itself as a foundation for elementary arithmetic. The idea is that the right-hand side of the biconditional embedded in Hume's Principle gives the truth conditions for the left-hand side of the biconditional, and the left-hand side has the grammatical and logical form it appears to have. In particular, locutions like "the number of *F*" are genuine singular terms, the linguistic forms used to denote objects. At least some instances of the right-hand side of Hume's Principle are true, on logical grounds alone. For example, it is a logical truth that the concept of "not identical to itself" is equinumerous with the concept "not identical to itself". Thus, from Hume's Principle, we conclude that the number of non-self-identical things is identical to the number of non-self-identical things. Letting "0" abbreviate "the number of non-self-identical things", we conclude that $0 = 0$, and so 0 exists.

Following Frege, the neo-logicist then defines the number 1 to be the number of the concept "identical to zero", the number 2 as the number of the concept "either identical to zero or identical to one", and on from there. It follows from HP that these natural numbers all exist and are different from each other.

No one doubts that Frege's theorem is a substantial mathematical achievement, showing how the natural number structure flows from a basic principle about cardinality. The philosophical significance of this is a matter of some discussion. What does Frege's theorem tell us about number, and about mathematics generally?

If Hume's Principle were a logical truth or, in Frege's terms, a "general logical law", then Frege's theorem would show that the basic principles of arithmetic were themselves logical truths—assuming that the principles of derivation used in Frege's theorem preserve this status. Clearly, however, HP is not a logical truth. It is not true in virtue of its form, nor does it seem to be derivable from accepted logical laws. Wright and Hale also do not call Hume's Principle a *definition* of (cardinal) number. As most logicians understand the notion, a definition must be *eliminable* in the sense that any formula containing the defined term is equivalent to a formula not containing it. It follows from Hume's Principle that there is something that is the number of non-self-identical things, in symbols $\exists x(x = 0)$. Hume's Principle does not provide for an equivalent sentence lacking the number terminology. A successful definition should also be *non-creative* in the sense that it has no consequences for the rest of the language and theory. Hume's Principle does have such consequences, since it entails that the universe is infinite. So Hume's Principle is neither eliminative nor non-creative.

Wright and Hale thus do not defend the traditional logicist thesis that arithmetic truth is a species of logical truth, nor that each arithmetic truth is true by definition. They argue instead that Hume's Principle is "analytic of" the concept of *natural number*. The claim is that we can account for the necessity of at least the basic arithmetic truths and how these truths can be known a priori. In a later work, Wright [36, pp. 210–211] wrote:

> Frege's theorem will... ensure... that the fundamental laws of arithmetic can be derived within a system of second-order logic augmented by a principle whose role is to *explain*, if not exactly to define, the general notion of identity of cardinal number, and that this explanation proceeds in terms of a notion which can be defined in terms of second-order logic. If such an explanatory principle... can be regarded as *analytic*, then that should suffice... to demonstrate the analyticity of arithmetic. Even if that term is found troubling,... it will remain that Hume's Principle—like any principle serving implicitly to define a certain concept—will be available without significant epistemological presupposition... So one clear a priori route into a recognition of the truth of... the fundamental laws of arithmetic... will have been made out. And if in addition [Hume's Principle] may be viewed as a *complete* explanation—as showing how the concept of cardinal number may be fully understood on a purely logical basis—then arithmetic will have been shown up by Hume's Principle ... as transcending logic only to the extent that it makes use of a *logical* abstraction principle—one [that] deploys only logical notions. So,... there will be an a priori route from a mastery of second-order logic to a full understanding and grasp of the truth of the fundamental laws of arithmetic. Such an epistemological route... would be an outcome still worth describing as logicism...

According to Alberto Coffa [7], a major item on the agenda of philosophy throughout the nineteenth century was to account for the necessity of mathematics and logic without invoking Kantian intuition. Coffa argued that the most successful line on this was what he called the *semantic tradition*, running through the work of Bernard Bolzano, Frege, Bertrand Russell, Ludwig Wittgenstein, David Hilbert, and culminating with Moritz Schlick and Rudolf Carnap in the Vienna Circle. With retrospect, the key idea is that the necessity of mathematics and logic lies in language, with meaning: a priori knowledge is knowledge of language use. The semantic tradition thus provided a line on the epistemology of mathematics and logic: we know mathematics to the extent that we know our own language. Of the various programs in the philosophy of mathematics alive today, Scottish neo-logicism is among the closest to the spirit of the semantic tradition.

Wright concedes that his own proposals hinge on the proviso that "concept-formation by abstraction" is accepted. That is, the program depends on the legitimacy of introducing at least some concepts via abstraction principles. Neil Tennant [34, p. 236] and George Boolos [4] argued against "concept-formation by abstraction" as a legitimate maneuver for a prospective logicist (or anybody else, for that matter). The most prevalent and influential of these arguments is the "bad company objection". The claim is there is no non-ad hoc way to distinguish good abstraction principles like Hume's Principle, from bad ones like Basic Law V. To be sure, Hume's Principle is consistent while Basic Law V is not, but that distinction is too coarse-grained. Hume's Principle is an "axiom of infinity" in the sense that it is satisfiable only in infinite domains. Boolos points out that there are consistent principles in the form (ABS) that are satisfiable only in *finite* domains. If Hume's Principle is acceptable, then so are these others. However, these finite principles are incompatible with Hume's Principle. How then to distinguish the legitimate abstraction principles? Wright's [36, 38] response is to delimit and defend certain conservation principles which rule out the "bad" abstraction principles and allow the good ones, Hume's Principle in particular. The debate continues, and we will briefly return to it below.

As significant as the abstractionist development of arithmetic may be, arithmetic is only a small part of mathematics. Another major item on the neo-logicist agenda is to extend the treatment to cover other areas of mathematics, like real analysis, functional analysis, and perhaps geometry and set theory. The program involves the search for abstraction principles rich enough to characterize more powerful mathematical theories, but still acceptable "without significant epistemological presupposition" (see Wright [36, pp. 233–244], Shapiro [27, 28], and Hale [16, 17]). Here we will stick to the status of Hume's principle as a foundation of arithmetic.

2 The Measures

Frege [13, Section 2] observes that "it is in the nature of mathematics to prefer proof, where proof is possible", noting that "Euclid gives proofs of many things which anyone would concede him without question". This observation is correct. Euclid, Archimedes, Cauchy, Weierstrass, Dedekind, Frege, and a host of others provide rigorous proofs of "many things that formerly passed as self-evident", as Frege put it. He sets himself the task of providing proofs of such basic arithmetic propositions as "every natural number has a successor", the induction principle, and simple arithmetic identities like "$1 + 1 = 2$".

Thinkers differ widely on *why* we prefer proof. And scholars differ on why *Frege* thought that we prefer proof when proof is possible. What were the aims of Frege's logicism? What did Frege take himself to have accomplished in his proofs of basic arithmetic truths?

Robin Jeshion [21, pp. 939–940] summarizes various aims that have been attributed to Frege by different scholars:

> **Mathematical Rationale**: Frege's motives are mathematical. He desired to prove some theorems. He believed that whatever admits of proof ought to be proved. And he thought that the propositions of arithmetic, hitherto unproved, admit of proof. So they should be proved.
> **Logico-Cartesian Rationale**: Frege is a reformer who aims to perfect arithmetical knowledge. He thought that the logical source alone is capable of producing knowledge possessing absolute self-evidence, certainty, and clarity. He also thought that actual arithmetical knowledge is marred by doubt, uncertainty, and unclarity. He aimed to establish logicism because he thought doing so was necessary for demonstrating the epistemological superiority of arithmetical knowledge.
> **Knowledge-of-Sources Rationale**: Frege desired to discern the philosophical status of our arithmetical knowledge i.e., to determine whether arithmetic is analytic or synthetic, a priori or a posteriori. He aimed to establish logicism because he thought that proving the propositions of arithmetic was necessary for determining the epistemological source of our arithmetical knowledge.

Jeshion observes that these rationales are compatible with each other. The exegetical issue in dispute is whether Frege had only one of these rationales, at least primarily. We need not enter into those matters here.

Jeshion delimits a fourth rationale, which she takes to be the correct or primary one. It is formulated in terms of some metaphysical cum epistemological principles concerning propositions. To articulate this rationale, we must take a small detour through Frege's rationalism.

In one place, Frege [13, Section 2] does tell us why it is that it is that mathematicians "prefer proof, where proof is possible", at least metaphorically:

> The aim of proof is, in fact, not merely to place the truth of the proposition beyond all doubt, but also to afford us insight into the dependence of truths upon one another. After we have convinced ourselves that a boulder is unmoveable,... there remains the further question, what is it that supports it so securely?

Frege thus believed that propositions have dependence relations to one another. The dependence relations are objective, in the sense that it is not a matter of how some person or other comes to believe a given proposition. Rather, some propositions rest on others, objectively.

Frege's account of the notions of analyticity and a priority are formulated in terms of these dependency relations:

> [T]hese distinctions between a priori and a posteriori, synthetic and analytic, concern, as I see it, not the content of the judgement but the justification for making the judgement. Where there is no justification, the possibility of drawing the distinctions vanishes. When... a proposition is called a posteriori or analytic in my sense, this is not a judgement about the conditions, psychological, physiological, and physical, which have made it possible to form the content of the proposition in our consciousness; nor is it a judgement about the way

> in which some other man has come... to believe it true; rather it is a judgement about the ultimate ground upon which rests the justification for holding it to be true...
>
> The problem becomes... that of finding the proof of the proposition, and of following it up right back to the primitive truths. If, in carrying out this process, we come only on general logical laws and on definitions, then the truth is an analytic one... If, however, it is impossible to give the proof without making use of truths which are not of a general logical nature, but belong to the sphere of some general science, then the proposition is a synthetic one. For a truth to be a posteriori, it must be impossible to construct a proof of it without including an appeal to facts, i.e., to truths which cannot be proved and are not general... But if, on the contrary, its proof can be derived exclusively from general laws, which themselves neither need nor admit of proof, then the truth is a priori. [13, Section 3]

Let us set aside the possibility that the dependency relation among propositions is not well-founded. It follows that some true propositions are foundationally secure and not grounded on other propositions. Frege's term for such propositions is *selbstverständlich*. Following Jeshion, I will leave this in the German. In the terminology of contemporary philosophy, "primitive", or "primitive truth", may do. *Selbstverständlich* propositions require no proof and, indeed, no (non-trivial) proof them is possible. All other known propositions are based on *selbstverständlich* propositions. So proper axioms are *selbstverständlich*.

How, then, are the *selbstverständlich* truths knowable? By definition, such propositions cannot be proved. They are not (properly, foundationally) known on the basis of anything else. Frege held that proper axioms have an epistemic property that he calls *einleuchten*, that of being self-evident. Jeshion [21, p. 953] glosses the property as follows:

> (S-E) A proposition p is *self-evident* if and only if clearly grasping p is [a] sufficient and compelling basis for recognition of p's truth.

According to Tyler Burge [6, p. 312], the epistemic-metaphysical status enjoyed by such propositions is "something like *beyond a reasonable doubt by someone who fully understands the relevant propositions*".

Jeshion takes the aims of Frege's logicism to be the following:

> **Euclidean rationale**: Frege thought that primitive truths of mathematics have two properties. (i) They are *selbstverständlich*: foundationally secure, yet are not grounded on any other truth, and, as such, do not stand in need of proof. (ii) And they are self-evident: clearly grasping them is a sufficient and compelling basis for recognizing their truth. He also thought that the relations of epistemic justification in science mirrors the natural ordering of truths: in particular, what is self-evident is *selbstverständlich*. Finding many propositions of arithmetic non-self-evident, Frege concluded that they stand in need of proof.

Our concern here, of course, is not with Frege's logicism, but with Scottish neologicism, the program of founding branches of mathematics on abstraction principles and, in particular, of founding arithmetic on Hume's principle. The plan is to see how well the program fares according to the four rationales suggested by Jeshion.

3 Mathematical Rationale

> Frege's motives are mathematical. He desired to prove some theorems. He believed that whatever admits of proof ought to be proved. And he thought that the propositions of arithmetic, hitherto unproved, admit of proof. So they should be proved.

The thesis here is that Frege saw that the principle that every number has a successor and the induction principle, to take two examples, had not yet been proved. For mathematical reasons alone—since mathematics demands proof when proof is possible—Frege set out to prove those propositions. This would account for his dissatisfaction with the derivation of those propositions from Hume's principle. It is not that there is something wrong with that derivation. Rather, perhaps, Frege thought that he discovered that Hume's principle itself could be proved, once the proper explicit definitions of the cardinal numbers is provided. So he proved Hume's principle from those definitions, and then showed how to prove the successor and induction principles from there.

Of course, Frege did not give this as the reason why he was not satisfied with starting with Hume's principle, and why he moved to the explicit definitions of numbers in terms of extensions. But we are not going to get embroiled in exegetical matters here, and our concern is not with Frege but with Scottish neo-logicism.

For what it is worth, Wright and Hale's own motives are explicitly philosophical. Consider, for example, the very title of Wright [36], "On the philosophical significance of Frege's theorem". Our focus here, however, is less on the motives of the founders and more on how well the program fares on various rationales it might have. We are asking what they accomplished, not so much what they intended to accomplish.

As noted, there is little doubt that Frege's theorem is a significant mathematical achievement. Who would have thought that so much could be derived from such a simple and more or less obvious fact about cardinality? This is not sufficient to satisfy the mathematical rationale, at least as it is formulated here. According to it, Frege felt that the basic propositions of arithmetic were hitherto unproved, and he provided what were, in effect, the first *proofs* of these propositions.

The analogous claim on behalf of Scottish neo-logicism is that the derivation of the Peano-Dedekind axioms from Hume's principle provides what are, in effect, proofs of those propositions, perhaps the first such proofs. Prior to the advent of neo-logicism, the propositions were just assumed, perhaps as axioms or as definitions.

There are deep philosophical issues that lie just below the surface. We cannot know if the mathematical rationale is satisfied unless we know what the content of the rationale is, and we cannot know that unless we know what mathematics is and, in particular, what proof is. More locally, we need to know what arithmetic is, and what it takes to prove something in arithmetic. Variations of these issues will occupy us with the other rationales, so we might as well broach them here.

Our first question is this: What constitutes a proof? Consider, for example, the proposition *S* that every natural number has a successor. Frege and the neo-logicist each provide a valid derivation whose last line has the same words as a sentence that expresses S. The neo-logicist derivation has an advantage over Frege's in that, arguably, it's premises are true. But this feature alone is surely not sufficient to constitute a proof. Deriving S from $S\&S$ is not a proof of S.

A second, perhaps more important question concerns the content of the propositions in the derivation of Frege's theorem. Is the last line in the neo-Fregean derivation the same as S, the proposition that every natural number has a successor? It would have to be if we are to claim that we have a proof *of* S. The logicist and neo-logicist should not change the subject.

To take a silly example, suppose that I define "natural number" to be "child", and define "successor" to be "parent". Then I prove, by analytic reflection, the sentence "every natural number has a successor". To take a less silly example, suppose that I define a "natural number" to be a finite, von Neumann ordinal, and define "the successor of a 'natural number'" x to be $x \cup \{x\}$. Then I prove, from the axioms of union and pairing, that "every natural number has a successor".

In neither case have we proved S, the proposition that every natural number has a successor. As the saying goes, calling a tail a leg does not make a tail into a leg.[1] Calling a child or a von Neumann ordinal a natural number does not make a child or a von Neumann ordinal a natural number. The first interpretation, of course, is a complete non-starter. What we have proved, if anything, is that every child has a parent. The second, von Neumann, interpretation at least has the advantage of getting the correct structure. That is, the natural numbers are isomorphic to the finite von Neumann ordinals.[2]

For Frege, the interpretation of the successor principle S is something like this:

> Let x be the extension of a concept of the form "equinumerous with C", and assume that only finitely many objects fall under C. Then there is an x' and a concept D such that (i) there is a z such that Dz and C is equinumerous to the concept "falling under D but different from z", and (ii) x' is the extension of the concept "equinumerous with D".

---

[1] The allusion here is to an old joke, attributed to Abraham Lincoln. Question: If you call a tail a leg, then how many legs does a dog have? Answer: Four. Calling the tail a leg does not make it a leg.

[2] A structuralist about mathematics, such as myself, claims that the natural numbers just are the places in the structure exemplified by the finite von Neumann ordinals [26, 24]. Isomorphic systems are equivalent, and so any statement in the language of arithmetic that is true of the finite von Neumann ordinals is thus true of the natural numbers. So, for a structuralist, one can prove things about the natural numbers by proving things about the finite von Neumann ordinals. According to the structuralist, the second interpreter above has not *completely* changed the subject. Let T be the proposition that if x is a finite von Neumann ordinal, then so is $x \cup \{x\}$. Our second interpreter provides a proof of T from some evident principles of set theory. Given the isomorphism, T and the successor principle S are equivalent. Even so, our interpreter cannot rightly claim to have proved S. The reason, of course, is that one must prove T in order to (justifiably) claim that the von Neumann ordinals exemplify the natural number structure. That is, our structuralist cannot exploit the isomorphism until she has shown that the isomorphism holds, and she must establish T in order to do this. The structuralist holds that S is part of the very characterization of the structure in question. Of course, it is not exactly news that neither Frege nor the Scottish neo-logicists are structuralists.

Call this U. Frege provides a derivation of U in his (inconsistent) theory of extensions. For this to count as a proof of S, on the mathematical rationale, U and S must have the same content. That is, U must *be* the successor principle S. And this holds only if Frege's explicit definitions are correct. That is, Frege provides a proof of the successor principle only if the natural numbers spoken of by mathematicians and ordinary folk, from antiquity until 1884 are, in fact, the extensions that Frege says they are. I do not know how to argue for, or against, this. It seems to be the issue of how one adjudicates claims of conceptual or logical analysis. We will encounter similar matters again.

It seems to me that the neo-logicist is in a stronger position here. No matter what one's views on conceptual analysis, it does seem plausible that natural numbers just *are* cardinal numbers of finite concepts—assuming that we can put aside the role of natural numbers as ordinals. Intuitively, zero just is the cardinal number of a concept under which nothing falls, one is the cardinal number of singleton concepts, etc. Moreover, Hume's principle is a fairly obvious and straightforward truth about cardinal numbers, or at least it is if one thinks that cardinal numbers exist and that every concept has a cardinal number. Frege's theorem includes the derivation of the successor principle S from this more or less obvious truth. It is not a stretch to think of this as a *proof* of S.

But things are not completely straightforward. Both Frege and the Scottish neo-logicist derive the structure of the natural numbers—the finite cardinal numbers—from HP. In a sense, Richard Dedekind [9] does the reverse. He defines a *simply infinite system* to be a set with the requisite structure, and then defines the "natural numbers" to be one such system. He then gives a procedure for assigning a natural number to each finite collection. A version of HP, restricted to finite concepts, is then forthcoming.[3] One might think, in line with the general mathematical rationale, that we now have a *proof* of this restricted version of HP, from structural principles about the natural numbers—including the successor principle S—and straightforward definitions.

Frege held that his account gave the correct definition of the natural numbers, and thus the correct proof of S. This is due to his insistence that the proper definition of mathematical objects should flow from their (main) applications. The underlying theme has been called "Frege's constraint" (adopted explicitly by Hale [16], and discussed by Wright [39]). As noted, Dedekind, along with the structuralist (see note 2 above) disagree, taking the structure to be the essence of the natural numbers. The application of the numbers, as cardinalities, is added later, and the relevant theorem—HP—is then proved. The parties thus disagree on the mathematical rationale. One side takes themselves to have proved S and the other structural features of the natural numbers, from HP (plus definitions); the other takes themselves to have proved a restricted version of HP from the structure (plus definitions).

The set theorist also gets into this act. She defines a *cardinal number* to be a von Neumann ordinal that is not equinumerous with a smaller von Neumann ordinal.

---

[3] This restricted version of HP is also sufficient to derive the Peano-Dedekind axioms. See [19].

She then gives a definition of the “number of” each set, and proves a version of HP restricted to sets.

Which of these is the real proof? Will the real (finite) cardinal numbers please stand up? There is no need to adjudicate this piece of metaphysics here. My feeling, for what it is worth, is that there is something wrong with the question—although perhaps I should not push my own structuralist agenda here.

Perhaps the neo-logicist can argue as follows: the account of arithmetic we get from HP has the right structure *and* it accounts for standard applications.[4] So the theory based on HP can play the role that arithmetic does in our conceptual scheme—in our form of life. Perhaps that is all that matters for the mathematical rationale. If it walks like a duck, sounds like a duck, and looks like a duck—and does all the work that ducks do—then, doggone it, it is a duck. The same goes for the Dedekind/structuralist, and set-theoretic accounts as well. They, too, deliver the right structure and the right account of cardinal applications. Perhaps the accounts need not be seen as competitors.[5] I take it that this, too, is a structuralist insight.

In any case, all parties to the metaphysical dispute can agree that a restricted version of HP, plus definitions, is *equivalent* to the Peano-Dedekind axioms, plus definitions. We have learned this from Frege, neo-logicists of all sorts, and their opponents. It seems reasonable, and uncontroversial, to count this as falling somewhere within the mathematical rationale.

4 Logico-Cartesian Rationale

> Frege is a reformer who aims to perfect arithmetical knowledge. He thought that the logical source alone is capable of producing knowledge possessing absolute self-evidence, certainty, and clarity. He also thought that actual arithmetical knowledge is marred by doubt, uncertainty, and unclarity. He aimed to establish logicism because he thought doing so was necessary for demonstrating the epistemological superiority of arithmetical knowledge.

From this perspective, the logicist seeks to shore up our knowledge of arithmetic, presupposing that there was something defective about our previous epistemic state. We did not know the basic truths of arithmetic with as much certainty as possible. This rationale suggests that before logicism, we did not really know *that* the boulder of arithmetic is unmoveable—or at least we did not know this as well as we could and should. The logicist derivations remove any doubt, uncertainty, and unclarity in the basic propositions of arithmetic.

A neo-logicist who takes on this rationale must argue, or at least claim, that Hume’s principle is itself *less* open to doubt and is *more* certain and more clear

---

[4] Thanks to Crispin Wright here.

[5] The question here is of a piece with the issue of whether, for example, the natural number 2 is the same as the real number 2 and the complex number 2. For a related issue, in the context of Scottish neo-logicism, see [8].

than such arithmetic truths as the successor and induction principles, and even basic sums as $7 + 5 = 12$. This strikes me as hard to maintain.

As noted in section 1 above, one of the most discussed objections to neo-logicism is Tennant's and Boolos's claim of "bad company". As noted, Hume's principle has the same form (ABS) as Basic Law V. So clearly not every abstraction principle is acceptable. How can we be that sure that HP is OK? We do have a lovely proof, due to Boolos [3] and others, that HP is consistent if second-order arithmetic is. But surely that cannot convince anyone that HP is *more* certain or is otherwise better grounded epistemically than arithmetic.

Note also that Hume's Principle has some content that goes beyond that of the natural numbers, and their applications (see Heck [19] and, again, Tennant [34, p. 236]). The quantifiers at the start of HP range over all concepts, not just those that apply to only finitely many things. HP thus entails that every concept has a number. So HP entails that there is a number of all sets, a number of all ordinals, and, indeed, a number of *all* objects. Tennant and Boolos each turned this observation into another objection against the neo-logicist claim that HP is all but analytic. No matter how this ongoing debate is adjudicated, it seems clear that one cannot seriously claim that HP is more certain than arithmetic. The extra content of HP engenders possible worries and doubts that have nothing to do with arithmetic.

If a Scottish neo-logicist wanted to maintain the Logico-Cartesian rationale, as it is articulated here, the best route would be to follow the advice in Heck [19] and replace Hume's Principle with a version whose opening quantifiers are restricted to concepts that apply to at most finitely many objects. Call this Finite-HP.

To state the restriction on the quantifiers of Finite-HP, we would need a rigorous definition of finitude that does not presuppose the natural numbers. Surely, it won't do to restrict the opening quantifiers to concepts whose numbers are natural numbers. Shapiro [25, Section 5.1.2] gives a definition of Dedekind-finite using a second-order language, but this principle is equivalent to finitude only if one assumes a version of the axiom of choice. Another definition is that a concept C is finite if every total ordering on C has a first and last element. The content of the various definitions, and their relation to the ordinary notion of finitude are far from obvious.

Let us put this matter aside, and assume that we have a decent, sharp, and clear formulation of finitude. Then, arguably, Finite-HP is not any less uncertain or open to unclarity and doubt than second-order arithmetic is. Even so, however, finite-HP is surely not *more* secure than second-order arithmetic. Finite-HP does not really increase our certainty of arithmetic, or provide a better epistemic foundation. The two theories are equi-consistent. Indeed, with some straightforward definitions, the two theories are all but notational variants of each other.

One natural way to understand the terms "self-evident", "certainty", "clarity", and "doubt", in the formulation of the Logico-Cartesian Rationale, is in their ordinary, psychological senses. Prima facie, certainty, clarity, and doubt are states of mind. Any self-respecting Fregean logicist would reject this rationale, so interpreted, along the lines of Frege's onslaught against psychologism. Logicism is not a matter of anyone's subjective state of mind concerning basic arithmetic propositions. As we noted above, Frege wrote that what is at stake is not "a judgement about they way in which some ... man has come ... to believe" a given arithmetic proposition. Rather, "it is a judgement about the ultimate ground upon which rests the justification for holding it to be true ..." [13, Section 3]. Also,

> we are concerned here not with the way in which [the laws of number] are discovered but with the kind of ground on which their proof rests; or in Leibniz's words, 'the question here is not one of the history of our discoveries, which is different in different men, but of the connection and natural order of truths, which is always the same' [13, Section 17], Leibniz, *Nouveaux Essais*, IV, Section 9

Interpreting the Logico-Cartesian rationale for Scottish neo-logicism in terms of more objective epistemological relations takes us to the two rationales to be considered next.

There is another sort of Logico-Cartesian rationale. Our neo-Fregean might hold that because our previous grasp of arithmetic is marred by doubt, uncertainty, and unclarity—however that is to be understood—we should just chuck the old concepts, and replace them with some that are clearer, better known, and can do all the legitimate work that our previous notions did. A Scottish neo-logicist who adopts this perspective takes herself to be proposing something like a Carnap-style *explication* or Quinean regimentation of arithmetic terminology.[6] She does not really care whether the new, defined theory is an exact match to the previous, pre-theoretic array of concepts and notions. All that matters is that the new theory do all the legitimate work that the old one did.

According to an orientation to the mathematical rationale broached at the very end of the previous section, all that matters is that the theory based on HP have the right structure and play the role that arithmetic does in our conceptual scheme. If it walks like a duck, and does all the work that ducks do, then it is a duck. The present orientation is similar in spirit, in that all that matters is the structure and the typical applications. The crucial point here is that the neo-logicist might not care whether the subject has changed—the new thing may not be a duck. Perhaps the subject needed to be changed, since the old stuff was simply no good, or at least not as good as the new stuff. Someone who wants to follow this route needs an argument that there is indeed something defective concerning our epistemic states about the old arithmetic and number theory. If it ain't broke, then why fix it? And we would need some account of the epistemic status of the new subject matter. Why is it any better? Better in what ways? This, too, takes us to the other rationales for logicism.

5 Knowledge-of-Sources Rationale

> Frege desired to discern the philosophical status of our arithmetical knowledge i.e., to determine whether arithmetic is analytic or synthetic, a priori or a posteriori. He aimed to establish logicism because he thought that proving the propositions of arithmetic was necessary for determining the epistemological source of our arithmetical knowledge.

---

[6] Tarski [31, 32] may have had something like this in mind for the notions of truth and logical consequence.

This rationale is explicitly endorsed by Wright who, as we saw above [36, pp. 210–211], argued that “Hume’s Principle [is] available without significant epistemological presupposition” and concluded, from this, that “one clear a priori route into a recognition of the truth of … the fundamental laws of arithmetic … [has] been made out”. The claim, then, is that the program shows that arithmetic is knowable a priori. Frege’s theorem provides an “a priori route from a mastery of second-order logic to a full understanding and grasp of the truth of the fundamental laws of arithmetic”.

Most of the ongoing debates concerning Scottish neo-logicism focus on this rationale. I cannot do justice here to all of the subtleties and insights in the literature. My more modest goal is to show how an explicit formulation of the Knowledge-of-Sources rationale allows us to demarcate the battle lines, and to determine the burden of proof on each side of this debate.

A defense of neo-logicism on the knowledge-of-sources rationale would consist of the following steps:

(1) articulating the special epistemic status that is claimed for arithmetic, which amounts to explaining the claimed epistemic source of the basic propositions of arithmetic.
(2) arguing that Hume’s principle, and other acceptable abstraction principles, have this special status.
(3) arguing that the axioms and rules of second-order logic invoked in Frege’s theorem preserve the epistemic status in question. That is, the Scottish neo-logicist must show that if the premises of an invoked rule of inference enjoy the requisite epistemic status, then so does the conclusion.
(4) arguing that the conclusions at the end of Frege’s theorem are indeed statements of the basic propositions of arithmetic, i.e., arguing that the subject has not changed.

As for item (1), the Scottish neo-logicist need not adopt Frege’s rationalist understanding of the notions of a priority and analyticity, in terms of objective grounding relations among propositions. On standard conceptions, a proposition is a priori if it can be known independently of experience. As Simon Blackburn [1, p. 21] puts it, a proposition is known a priori if the knowledge is not based on any “experience with the specific course of events of the actual world”. Experience may be required to grasp the relevant concepts, but beyond that, no particular experience is necessary to know that the proposition is true. Surely, a central claim of the Scottish neo-logicist program is that the basic propositions of arithmetic are knowable a priori, in this sense.

I take it that on the contemporary understanding of the term, a sentence or proposition is analytic if it is true solely in virtue of the meaning of the sentence or proposition. We saw that the Scottish neo-logicists back off from the claim that the propositions of arithmetic are analytic, in this sense (section 1 above). But they do argue that Frege’s theorem shows that the basic truths of arithmetic can become known “without significant epistemological presupposition”. In particular, the basic truths can become known without an appeal to Kantian intuition, insight into a Platonic realm, empirical confirmation, indispensability to scientific practice, or the like. Let us focus on this, somewhat modest goal. Note that the neo-logicist does *not* claim that the program shows arithmetic to be known infallibly.There may indeed be some risk involved in accepting Hume's principle.

On (2): The main question in the extensive debate over Scottish neo-logicism is whether Hume's principle, or perhaps a restricted version of it, is indeed knowable without significant epistemological presupposition. I noted above, several times, that Hume's principle (or Finite-HP) is a fairly obvious truth, provided one assumes that cardinal numbers exist and that every concept (or that every finite concept) has a cardinal number. There, of course, is the rub—or at least a rub. Do we know that cardinal numbers exist and that every (finite) concept has a cardinal number, a priori, without significant epistemological presupposition? Wright and Hale say that we do, since we can derive those things from HP. But this takes us in a circle.

To avoid begging any questions, we should use a free logic to interpret the reasoning behind Scottish neo-logicism. Well-formed terms in the form "the number of F" should not *automatically* denote objects, cardinal numbers in this case. That is, we don't want to get the existence of numbers from the mere fact that expressions in the form "the number of F" are grammatical singular terms. To stay as close as possible to the Scottish neo-logicists' texts, the most natural free logic system to adopt is one in which an identity $s = t$ implies that s and t both exist, since Wright and Hale infer the existence of zero from the fact that $0 = 0$ follows from HP. It is not obvious that Hume's principle itself, so construed, has the requisite epistemic status, of being knowable without significant epistemological presupposition, just because HP has ontological consequences. That is, we derive the existence of zero from HP and the logical truth that the concept of being non-self-identical is equinumerous with itself. I have little to add here to this ongoing debate.

The thesis of Wright and Hale [40] is that Hume's principle is (or is akin to) an implicit definition of the concept of cardinal number, and is the sort of thing that can be true by stipulation. We, or the community of neo-logicists, lay it down as defining the concept of cardinal number. The claim is that HP is all but analytic, in the above sense—true in virtue of the stipulated meaning of the concept of "cardinal number".

Of course, not every principle in the form (ABS) successfully confers a meaning on its constituent concepts. There is still the bad company objection to deal with (again, see section 1 above). Wright and Hale have articulated some constraints that a successful (abstractionist) stipulation must meet. The details are not important, and they are subject to revision in light of further reflection anyway. Call the correct constraints C. The thesis is that if a principle A in the form (ABS) meets C, then A can be laid down as a giving the meaning of the operator on its left hand side. Let us assume that HP itself meets C.

A nice question now concerns what the epistemic status of C should be, in order for a stipulation, or an implicit definition, to be successful, and confer the requisite meaning on the key terms. It is surely too much to demand that the neo-logicist *prove* that his abstraction meets C before he can claim that it determines a meaning for its operator. Dialetheism notwithstanding, one necessary condition on successful abstractions is that they be consistent. The inconsistency of Basic Law V got us going on bad company in the first place. So, plausibly, consistency (or something that entails consistency) is among the constraints C. Gödel's second incompleteness theorem makes it unlikely, perhaps impossible, that one can prove the consistency of HP or Finite-HP without assuming something whose consistency is at least as problematic (such as the consistency of another theory or transfinite induction up to a large, countable ordinal). Can we, in fact, come to know the consistency of HP or Finite-HP independently of our firm conviction in, say, the Peano postulates? Do we have to? Perhaps the neo-logicist is entitled to just assume that the appropriate condition C holds, in a given case, so long as there is no evidence to the contrary. The idea is that an abstraction can be assumed to be innocent until proven guilty. Of course, if a given abstraction, such as HP, does not meet C, then its stipulation fails, perhaps unbeknownst to the neo-logicist. But, as noted above, we are not looking for infallibility, just a priority and lack of significant epistemological presupposition. The debate continues.[7]

Item (3) in the above list is the claim that the axioms and rules of second-order logic invoked in Frege's theorem preserve the requisite epistemic status. The axioms and rules in the neo-logicist derivation are, first off, the usual introduction and elimination rules for the connectives, first-order quantifiers, and second-order quantifiers. I presume that there is no serious question about those, once we agree on a free logic. The only other item invoked in the derivation is the comprehension scheme:

$$\exists X \forall x (Xx \equiv \Phi(x)),$$

one instance for each formula Φ in the language that does not contain the variable X free.

Comprehension says, in effect, that there is a concept (or property, or set, or whatever it is that second-order variables range over) corresponding to each formula in the language. To derive Frege's theorem, one must invoke some *impredicative* instances of the comprehension scheme. That is, we need concepts corresponding to formulas that have variables that range over all concepts.[8] For example, the Fregean and neo-logicist definition of the successor relation has an existential quantifier ranging over concepts. So to prove things about this relation—that it is a function, is one-to-one, etc.—we need the existence of concepts obtained via comprehension over a formula with a bound variable. One key item, of course, is the induction principle, which is itself impredicative. We also need an instance of comprehension

[7] Philip Ebert and I pursue these issues further in "The good, the bad and the ugly" in a forthcoming special issue of *Synthese*, edited by Øystein Linnebo, devoted to the bad company objection.

[8] If we are allowed to introduce new predicate symbols along the way, Frege's theorem itself does not invoke any instance of comprehension in which the formula is more complex than Π^1_1. Of course, if we limit ourselves to Π^1_1 comprehension, then we can only derive induction for Π^1_1, formulas. There are other ways of formulating the deductive system, but parallel questions will arise in each case. It should be noted that there is a second type of impredicativity in Frege's theorem: the "number of" operator applies to concepts characterized with, or defined by, that very operator. See [22] for an extended discussion of the various uses of impredicativity.

in which the embedded formula is something like "$x \neq x$". That is, Frege's theorem requires the existence of an empty concept. There are sources of worry on these issues, and there is an extensive literature on the matters (see, for example, [34, 10, 37, 30]).

The final item (4) is the claim that the conclusions at the end of Frege's theorem are indeed statements of the basic propositions of arithmetic. We start with an abstraction, Hume's Principle, which is the sort of thing that can be stipulated to be true, and true because it is so stipulated. The question before us now is whether the neo-logicist has changed the subject. Are the conclusions of Frege's theorem indeed truths about the natural numbers that everyone invokes when doing pure and applied arithmetic?

We have dealt with a version of this issue, in section 3 above, with the mathematical rationale. Recall also that at the end of the previous section, we broached an interpretation in which the neo-logicist does not care whether the subject has changed, or even admits that the subject has changed. That orientation can be invoked here as well, in which case, item (4) comes off the agenda. But let us assume, for the present, that the neo-logicist is attempting an epistemological foundation for the arithmetic we already know and love.

One issue here concerns the nature of the "stipulation" claimed by the Scottish neo-logicist. The word "stipulation" at least suggests that the neo-logicist is intending to introduce a *new* concept, one that is not already in use. One cannot just stipulate truths about concepts as deployed in everyday and mathematical and scientific practice. I cannot stipulate, for example, that a tail is a leg, or that a natural number is a child or a finite von Neumann ordinal.[9] At best, this would introduce an ambiguity. There would be multiple uses of the word "tail" and "natural number", just as there are multiple uses of the word "bank".

Suppose, first, that we do understand the Scottish neo-logicist's stipulation as the introduction of a new concept. And suppose the stipulation is successful, as above. To avoid ambiguity, call the defined concept "SNL-number", and call the SNL-number of a finite concept a "SNL-natural-number". Then Frege's theorem, plus standard meta-theoretical results, entails that the SNL-natural-numbers do indeed exemplify the natural number structure. They are isomorphic to the natural numbers themselves, the finite von Neumann ordinals, etc. So the SNL-natural-numbers will do as surrogates for the natural numbers. Anything, in the language of arithmetic, that is true of them is true of any other system that exemplifies the structure. And all of the usual applications of the numbers are forthcoming. The structuralist in me is satisfied that the subject has not changed, no more than it has with the set-theoretic definitions. Again, this broaches a question of how concepts are to be individuated, a question that is at least of a piece with the Caesar problem (and the "C-R" problem broached in Cook and Ebert [8]).

The Scottish neo-logicist need not think of things as going this way. She may not take the "stipulation" of HP (or finite-HP) to be introducing a new concept, for

[9] Contrary to what I proposed in [26, p. 81]; see [29] for a correction.

the first time. If there was such a stipulation, it happened in antiquity, possibly in Greece, and this stipulation governs the current use of the notion of cardinal number.

In any case, to avoid the charge of changing the subject, our Scottish neo-logicist might drop the talk of stipulation, except perhaps as a rational reconstruction of our concepts. One possible thesis is that Hume's Principle is somehow implicit in our practice using the notion of (finite) cardinal number. To be frank, I have little idea of how one adjudicates claims like this. Is it an empirical matter about a *word*, to be settled by a field linguist studying native speakers of English or German, or is it an a priori claim about a concept, settled by conceptual analysis? The latter falls within Frege's own methodology (see [2]), but there is no real consensus of what the rules of this game are. I propose to leave the topic with this indecision.

6 Euclidean Rationale

> Frege thought that primitive truths of mathematics have two properties. (i) They are *selbstverständlich*: foundationally secure, yet are not grounded on any other truth, and, as such, do not stand in need of proof. (ii) And they are self-evident: clearly grasping them is a sufficient and compelling basis for recognizing their truth. He also thought that the relations of epistemic justification in science mirrors the natural ordering of truths: in particular, what is self-evident is *selbstverständlich*. Finding many propositions of arithmetic non-self-evident, Frege concluded that they stand in need of proof.

From Frege's perspective, once again, the properties of being *selbstverständlich* and self-evident are both *objective* features of propositions. The former concerns the natural order of truths, and the latter concerns the way in which some truths can become known, and perhaps should become known, reflecting the objective grounding relations.

A Scottish neo-logicist who accepts this rationale must show (1) that Hume's Principle is *selbstverständlich*, i.e., foundationally secure, and not grounded on any other truth in this objective sense of grounding, and (2) that HP is self-evident: clearly grasping it is a sufficient basis for knowing it, or at least recognizing its truth. Also, once again, there may be a matter of showing that the neo-logicist has not changed the subject—that HP is indeed about the cardinality operator in use all along.

If Hume's Principle is akin to an implicit definition, introducing a new concept, and is the sort of thing that can be true by stipulation, then, one would think, it is *selbstverständlich*. That is to say, if the concept of "cardinal number of" is in fact *introduced* by the abstraction, then there is no sense of proving HP from other principles. Indeed, if we prove HP from other principles, then, clearly, the notion of "cardinal number" was introduced some other way. However, as we saw in the previous section, if we think of the abstraction as a stipulation, then HP cannot be about the natural numbers we have had since antiquity, unless the stipulation, or something like that, took place then. The Scottish neo-logicist has changed the subject, if that matters.

If we are to think of HP as being about the same natural numbers we have used and talked about all along—and if we buy into the Fregean metaphysical cum epistemological framework—then there is a substantive issue of whether HP is *selbstverständlich*. Perhaps it is grounded in other principles. Apparently, Frege himself thought that HP is grounded in Basic Law V, plus the explicit definitions. As we saw, Finite-HP—or at least something expressed with the same words as Finite-HP—can be derived from second-order Peano-Dedekind arithmetic, plus straightforward definitions, ala Dedekind [9]; and a version of HP, restricted to sets, can be derived in ZF, plus straightforward definitions. The Scottish neo-logicist who adopts the Euclidean rationale must contend that these are not really proofs, or at least that HP does not *need* those proofs. The thesis is that HP is not grounded on the Peano-Dedekind or ZF axioms nor, of course, on an inconsistent theory of extensions. Concerning the Peano-Dedekind axioms, it is the other way around: the successor principle and the like are grounded on HP. Frankly, I do not see how this dispute is supposed to be adjudicated, or even what counts as an argument for or against the relevant theses. There are two sets of principles, and a version of each is derivable from a version of the other. We are given little guidance on how one can tell which of them grounds the other in the natural ordering of truths.

The second thesis, under the Euclidian rationale, is that Hume's principle is self-evident, in the Fregean sense. The claim is that clearly grasping HP is a sufficient and compelling basis for recognition of its truth. To use Burge's [6, p. 312] formulation, the claim is that HP is beyond a reasonable doubt for anyone who fully understands it.

Burge and Jeshion both allow the possibility that a given proposition can be self-evident in the requisite sense and yet open to rational doubt. They both discuss Frege's attitude toward Basic Law V. Even before the fateful letter from Russell arrived, Frege admitted that Basic Law V is not as clear, or as obvious, as one would like. As Burge [6, p. 337] put it, since Frege thought that Basic Law V was an axiom, "he must have, at least sometimes, thought that *it* was certain, but because of insufficient analysis or incomplete understanding, *he* was not".

The talk of "clearly grasping" and "fully understanding" theses like HP is of a piece with the rationalist conception of clear and distinct ideas. Accordingly, one can have a "confused" understanding of the concepts in various propositions. The thesis at hand is that once all of these "confusions" are cleared up, and one has a proper—clear and distinct—grasp of the concepts, one will know that HP is true.

To take the contrapositive of the thesis at hand, the idea is that if a proposition is self-evident, then anyone who has a reasonable doubt about it does not fully understand it. At best, she has only a confused grasp of the concepts in the proposition. For example, a nominalist, who rejects HP, does not really understand it. She does not fully know what she is talking about.

As Frege deploys the notion of self-evidence, it is crucial that the sufficient and compelling basis one gets for knowing the proposition in question does not include any deduction. Surely Frege thought that a full grasp of the concepts involved is sufficient to know sums like $135 + 784 = 919$ or even $7 + 5 = 12$ or $2 + 2 = 4$. After all, he thought that these sums are analytic. However, Frege explicitly argued that the sums are not self-evident. They are only known via deduction—we have to prove them. This disqualifies the sums as a candidates for self-evidence.

As we saw, the Scottish neo-logicist argues that HP is akin to an implicit definition; it specifies the meaning of the "number of" operator. Recall, again, that not every sentence in the form (ABS) is a successful implicit definition. There is the bad company objection to deal with. The neo-logicist specifies some constraints, which we label C, on a successful abstraction. The view is that an abstraction is good, and can become known by stipulation, only if C holds of it. In the previous section, we broached the question of the required epistemic status of the proposition that C holds of the abstraction.

Suppose that the orientation is that someone properly knows an abstraction only if she *proves* that C holds for it. Then HP is neither *selbstverständlich* nor self-evident. Understanding the concepts is not sufficient to know the truth of HP. One must also know that the constraints C hold of it. But, as noted in the previous section, it is surely too much to demand that someone prove that C holds in order to know that HP is true.

Suppose, instead, that the proper orientation is that abstractions like HP are innocent until proven guilty. In other words, one can just assume, or is entitled to assume, that the constraints C are met unless there is some reason to think otherwise. Even then, one might think, HP is not self-evident and not *selbstverständlich*. Presumably, one must at least know that there is no reason to doubt that C holds before one knows HP. To be sure, this extra knowledge is not a premise in the derivation of HP, but, still, such (negative) knowledge is required in order to know HP. Understanding alone, no matter how clear, is not sufficient.

The neo-logicist can go completely externalist here. The thesis is that an abstraction is good, and known, so long as the constraints C are met, whether anyone knows this, or wonders about this, or is blissfully ignorant of it. In this externalist framework, perhaps, HP *is* self-evident. If the constraints in question are, in fact, met, then understanding the proposition is to know it. Of course, one can then wonder what role the notion of self-evidence is to play in this externalist framework.

There is one more way to look at things. What are we to say of abstractions, like Basic Law V, for which the constraints are not met? There are (at least) two options. One is that Basic Law V does indeed confer a meaning on its embedded term, the extension operator. It is just that nothing answers to that meaning. Basic Law V is thus false, or is indefinite, due to (something analogous to) reference failure.[10] Another option is to insist that if the constraints fail, then *no* meaning is conferred. Recall

[10] Philip Ebert [11, Chapter 3] defines a stipulation to be "effective" if it confers a meaning on its previously undefined terms, and he defines a stipulation to be "successful" if, roughly, it is both effective and true. Suppose that someone says, "Let 'TC' be my oldest brother". Call this statement T. Prima facie, the stipulation of T does confer a meaning on the previously undefined term "TC". We now know how to deploy the term in conversation, what follows from typical uses of it, and the like. So the stipulation is effective, in Ebert's sense. But if the speaker has no brothers, then T fails to be true (due to reference failure), and so the stipulation was not successful.

that we are thinking of HP and Basic Law V as attempts to stipulate meanings for operators. The present option is that if the constraints on successful stipulation fail, then so does the attempt to stipulate a meaning. That is, no meaning is stipulated. Basic Law V, so construed, is literally meaningless.

If we go with the latter route, then HP is arguably self-evident. That is, HP is perhaps beyond a reasonable doubt for *anyone who fully understands it*. If somebody does in fact understand HP, then there is something to understand. In other words, if someone understands HP, then its terms are meaningful, and so the stipulation was successful. In that case, HP is true. Our subject may harbor a rational doubt as to whether she understands HP (if she has doubts about whether C is satisfied, for example). But, arguably, if she understands it, then she knows it.

From this perspective, the conditions C are more like presuppositions on the use of an implicit definition. One need not know that presuppositions are met in order to successfully deploy a concept or, arguably, to know things expressed with the concept. As with the previous option, we might think of the constraints as part of an externalist account of foundational knowledge. If the constraints are met, then we know that the abstraction is true. We need have no internal access or warrant for the truth of the constraints.

We have come to the end of the brief tour of some rationales for logicist-type programs in the foundations of mathematics. This is not the place to settle the delicate issues concerning this foundationalist orientation to mathematical knowledge, much less the extensive debates in the literature. I will rest content if I have helped to focus the discussion, both on the various rationales and the presuppositions of each.

Acknowledgments Thanks to Philip Ebert and Crispin Wright for providing valuable comments on a previous version of this paper. I much appreciate the collegiality.

References

1. Blackburn, S. [1994], *The Oxford dictionary of philosophy*, Oxford, Oxford University Press.
2. Blanchette, P. [1996], "Frege and Hilbert on consistency", *Journal of Philosophy 93*, 317–336.
3. Boolos, G. [1987], "The consistency of Frege's *Foundations of arithmetic*" in *On being and saying: Essays for Richard Cartwright*, edited by Judith Jarvis Thompson, Cambridge, Massachusetts, The MIT Press, 3–20; reprinted in [5], 183–201.
4. Boolos, G. [1997], "Is Hume's principle analytic?", in *Language, thought, and logic*, edited by Richard G. Heck, Jr., New York, Oxford University Press, 245–261; reprinted in [5].
5. Boolos, G. [1998], *Logic, logic, and logic*, Cambridge, Massachusetts, Harvard University Press.
6. Burge, T. [1998], "Frege on knowing the foundation", *Mind 107*, 305–347.
7. Coffa, A. [1991], *The semantic tradition from Kant to Carnap*, Cambridge, Cambridge University Press.
8. Cook, R. and P. Ebert [2005], "Abstraction and identity", *Dialectica 59*, 1–19.

9. Dedekind, R. [1888], *Was sind und was sollen die Zahlen?*, Brunswick, Vieweg; translated as *The nature and meaning of numbers*, in *Essays on the theory of numbers*, edited by W. W. Beman, New York, Dover Press, 1963, 31–115.
10. Dummett, M. [1991], *Frege: Philosophy of mathematics*, Cambridge, Massachusetts, Harvard University Press.
11. Ebert, P. [2005], *The context principle and implicit definitions: Towards an account of our a priori knowledge of arithmetic*, Ph.D. thesis, University of St. Andrews.
12. Fine, K. [2002], *The limits of abstraction*, Oxford, Oxford University Press.
13. Frege, G. [1884], *Die Grundlagen der Arithmetik*, Breslau, Koebner; *The foundations of arithmetic*, translated by J. Austin, second edition, New York, Harper, 1960.
14. Frege, G. [1893], *Grundgesetze der Arithmetik 1*, Olms, Hildescheim.
15. Hale, B. [1987], *Abstract objects*, Oxford, Basil Blackwell.
16. Hale, B. [2000a], "Reals by abstraction", *Philosophia Mathematica (3) 8*, 100–123; reprinted in [18], 399–420.
17. Hale, B. [2000b], "Abstraction and set theory", *Notre Dame Journal of Formal Logic 41*, 379–398.
18. Hale, B. and Wright, C. [2001], *The reason's proper study*, Oxford, Oxford University Press.
19. Heck, R. [1997], "Finitude and Hume's principle", *Journal of Philosophical Logic 26*, 589–617.
20. Hodes, H. [1984], "Logicism and the ontological commitments of arithmetic", *Journal of Philosophy 81*, 123–149.
21. Jeshion, R. [2001], "Frege's notions of self-evidence", *Mind 110*, 937–976.
22. Linnebo, Ø. [2004], "Predicative fragments of Frege arithmetic", *Bulletin of Symbolic Logic 10*, 153–174.
23. Parsons, C. [1964], "Frege's theory of number", *Philosophy in America*, edited by Max Black, Ithaca, New York, Cornell University Press, 180–203; reprinted in *Mathematics in Philosophy*, by C. Parsons, Ithaca, New York, Cornell University Press, 1983, 150–175.
24. Resnik, M. [1997], *Mathematics as a science of patterns*, Oxford, Oxford University Press.
25. Shapiro, S. [1991], *Foundations without foundationalism: A case for second-order logic*, Oxford, Oxford University Press.
26. Shapiro, S. [1997], *Philosophy of mathematics: Structure and ontology*, New York, Oxford University Press.
27. Shapiro, S. [2000], "Frege meets Dedekind: A neo-logicist treatment of real analysis", *Notre Dame Journal of Formal Logic 41*, 335–364.
28. Shapiro, S. [2003], "Prolegomenon to any future neo-logicist set theory: Abstraction and indefinite extensibility", *British Journal for the Philosophy of Science 54*, 59–91.
29. Shapiro, S. [2006], "Structure and identity", in *Identity and modality*, edited by Fraser MacBride, Oxford, Oxford University Press, 164–173.
30. Shapiro, S. and A. Weir [2000], "'Neo-logicist' logic is not epistemically innocent", *Philosophia Mathematica (3) 8*, 163–189.
31. Tarski, A. [1933], "Der Warheitsbegriff in dem formalisierten Sprachen", *Studia philosophica 1*, 261–405; translated as "The concept of truth in formalized languages", [33], 152–278.
32. Tarski, A. [1935], "On the concept of logical consequence", in [33], 417–429.
33. Tarski, A. [1956], *Logic, semantics and metamathematics*, Oxford, Clarendon Press; second edition, edited by John Corcoran, Indianapolis, Hackett Publishing Company, 1983.
34. Tennant, N. [1987], *Anti-realism and logic*, Oxford, Oxford University Press.
35. Wright, C. [1983], *Frege's conception of numbers as objects*, Aberdeen, Aberdeen University Press.
36. Wright, C. [1997], "On the philosophical significance of Frege's theorem", *Language, thought, and logic*, edited by Richard Heck, Jr., Oxford, Oxford University Press, 201–244; reprinted in [18], 272–306.
37. Wright, C. [1998], "On the harmless impredicativity of $N^{=}$ (Hume's principle)", *The philosophy of mathematics today*, edited by Mathias Schirn, Oxford, Oxford University Press, 339–368; reprinted in [18], 229–255.

38. Wright, C. [1999], "Is Hume's Principle analytic", *Notre Dame Journal of Formal Logic 40*, 6–30; reprinted in [18], 307–332.
39. Wright, C. [2000], "Neo-Fregean foundations for real analysis: Some reflections on Frege's constraint", *Notre Dame Journal of Formal Logic 41*, 317–334.
40. Wright, C. and Bob Hale [2000], "Implicit definition and the a priori", in *New essays on the a priori*, edited by P. Boghossian and C. Peacocke, Oxford, Oxford University Press, 286–319; reprinted in [18], 117–150.

Natural Logicism via the Logic of Orderly Pairing

Neil Tennant*

Abstract The aim here is to describe how to complete the constructive logicist program, in the author's book *Anti-Realism and Logic*, of deriving all the Peano–Dedekind postulates for arithmetic within a theory of natural numbers that also accounts for their applicability in counting finite collections of objects. The axioms still to be derived are those for addition and multiplication. Frege did not derive them in a fully explicit, conceptually illuminating way. Nor has any neo-Fregean done so.

These outstanding axioms need to be derived in a way fully in keeping with the spirit and the letter of Frege's logicism and his doctrine of definition. To that end this study develops a logic, in the Gentzen-Prawitz style of natural deduction, for the operation of orderly pairing. The logic is an extension of free first-order logic with identity. Orderly pairing is treated as a primitive. No notion of set is presupposed, nor any set-theoretic notion of membership. The formation of ordered pairs, and the two projection operations yielding their left and right coordinates, form a coeval family of logical notions. The challenge is to furnish them with introduction and elimination rules that capture their exact meanings, and no more.

Orderly pairing as a logical primitive is then used in order to introduce addition and multiplication in a conceptually satisfying way within a constructive logicist theory of the natural numbers. Because of its reliance, throughout, on sense-constituting rules of natural deduction, the completed account can be described as 'natural logicism'.

N. Tennant (✉)
Department of Philosophy, The Ohio State University, Columbus, Ohio 43210
e-mail: tennant.9@osu.edu

*Earlier versions of this paper were presented to the Arché Abstraction Weekend on Tennant's Philosophy of Mathematics in St. Andrews in May 2004 and to the Midwest Workshop in Philosophy of Mathematics at Notre Dame in October 2005. The author is grateful for helpful comments from members of both audiences, especially the St. Andrews commentators Peter Milne, Ian Rumfitt, Peter Smith and Alan Weir; and from George Schumm, Timothy Smiley and Matthias Wille. Comments by two anonymous referees have also led to significant improvements.

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_5, © Springer Science+Business Media B.V. 2009

1 Introduction: Historical Background

Frege's two-volume work *Grundgesetze der Arithmetik* ([5, 6]) fell significantly short of completing his logicist program for arithmetic. This is not because of the all-too-obvious reason that his underlying theory of classes fell prey to Russell's Paradox. Although the following counterfactual enjoys no cognitive sense (since its antecedent is impossible), it nevertheless makes good rhetorical sense:

> *Even if* Frege's system of classes had *not* turned out to be inconsistent, the technical work accomplished in the two volumes of *Grundgesetze* nevertheless cannot be judged to have achieved what Frege himself demanded of a logicist account of arithmetic.

In addition, then, to the criticisms of tone, arrangement and dialectical strategy that Dummett ([4], Chapter 19) has levelled against Part III (in Vol. II) of *Grundgesetze*, one can add the criticism that Frege did not thoroughly finish off a logicist derivation and justification of the basic laws of arithmetic.

Frege's goals were twofold:

1. The account should explain the general *applicability* of arithmetic, and
2. The account should afford fully formal deductions, from its underlying logicist principles, of the basic laws of arithmetic itself.

These basic laws have standardly been taken to include *at least* the Peano–Dedekind axioms for zero and successor:

$$\forall x\ \neg sx = 0;$$
$$\forall x \forall y\ (sx = sy \rightarrow x = y);$$

and the principle of mathematical induction:

$$\forall\Phi(\forall x(\Phi x \rightarrow \Phi sx) \rightarrow (\Phi 0 \rightarrow \forall y \Phi y)).^1$$

Frege had set himself the goal of furnishing a logicist derivation of at least these, from the time (1884) of his earlier work *Die Grundlagen der Arithmetik* onward. But we must not lose sight of the fact that the later, more technical *Grundgesetze* (Vol. I: 1893; Vol. II: 1903) appeared after the Peano–Dedekind recursion axioms for addition:

$$\forall x\ (x + 0 = x);$$
$$\forall x \forall y\ (x + sy = s(x + y))$$

and for multiplication:

[1] These axioms are stated without any explicit sortal restriction to the natural numbers. The quantifiers are to be understood as ranging over the natural numbers.

$$\forall x\ (x \times 0 = 0);$$
$$\forall x \forall y\ (x \times sy = (x \times y) + x)$$

had been established as part of the axiomatic basis for arithmetic. Peano's axioms were published in 1889; Dedekind's *Was sind und was sollen die Zahlen?*, with its justification of definition by recursion, appeared in 1888. But Frege, in his 1893 pursuit of the goal (1), could not justifiably omit or exclude the operations of addition and multiplication from the scope of logicism's obligations with regard to goal (1).

One possible explanation for the widespread belief that Frege had provided a logicist derivation *and thereby* a justification of all the laws of arithmetic is that Montgomery Furth, in his Introduction to his translation of introductory portions of the first volume of *Grundgesetze*, wrote ([7], p. vi) that '[f]or Frege, the paramount part of [his] task' was to 'actually produce derivations of the standard propositions of arithmetic ... '. This leaves the reader who relies on English translations of Frege's work with the impression that Frege did indeed furnish definitions and proofs (albeit within his inconsistent system) for *all* the basic laws of arithmetic, including those governing addition and multiplication—definitions and proofs, moreover, that would have enabled Frege to claim that goal (1) had thereby been accomplished.

Michael Potter has correctly observed ([13], at p. 75), concerning the *Grundlagen*, that 'What is missing is a *justification* of the recursive definitions of addition and multiplication' [emphasis added], and goes on to say that

> ... the treatment of the *definition* of functions by recursion is in fact the least obvious part of the deduction of arithmetic from Peano's axioms: ... it is one of the most impressive achievements of Dedekind's later [1888] work. So perhaps the reason for Frege's sudden silence at this point is that he had by then seen the difficulty but not yet worked out the solution. [Loc. cit.; emphasis and date added.]

Potter is referring to a 'sudden silence' at a certain point in Frege's exposition within the *Grundlagen.* As Potter says (ibid., at p. 79)

> ... there remained to Frege at the technical level the task, carried out fully in his *Grundgesetze* ten years later, of filling in the gaps.

One of those gaps, it is clear from Potter's discussion, he (Potter) takes to be 'the crucial step of *justifying* the recursive definitions of addition and multiplication'. [Emphasis added.] In the context of our current distinction between goal (1) of applicability and goal (2) of derivability, we should speak here on Potter's behalf of 'the crucial step of *deriving* the recursive *axioms for* addition and multiplication'. Only the latter, derivational, step of the project, but not the former, justificatory, one, was so much as addressed in *Grundgesetze*, let alone 'carried out fully'. There is room for complaint concerning how fully and explicitly this was actually done. The justificatory step, however, was not carried out *at all*. This surprising and disappointing discovery is what a close reading of *Grundgesetze* will reveal.

Richard Heck [10] argues that (at least in the case of addition) one can read into various of Frege's definitions and numbered theorems something tantamount to a formal derivation, by appeal to those definitions, of something like the Dedekind–Peano recursion axioms. Such an accomplishment, on Frege's part, would not be undermined by the subsequent discovery that his system is inconsistent. For, as Heck has also pointed out ([9], at p. 580)

> ... it has in recent years been shown that the Dedekind–Peano axioms for arithmetic can indeed be derived, within second-order logic, from Hume's Principle.

Heck's later paper [10] gives an exposition, in modern logical notation, of analogues of Frege's second-order definitions and theorems in this connection. Heck's analogues eschew Frege's own use of ordered pairs (hostage, with hindsight, to the misfortune of the inconsistency of Axiom V), and use instead what Heck calls the 2-ancestral. But the overall pattern of logical relations among the formal sentences involved is claimed to be preserved. Even so, Heck's charitably reconstructed Frege does not deal with multiplication at all.

Let us for the present confine ourselves to addition. There is warrant for the complaint that the actual derivations that Frege gave (and analogues of which Heck expounds) of the 'recursive definition' of addition do not amount to the *justification* that the Fregean should be demanding of a satisfactory logicist account of addition. Frege failed to complete the *philosophical* and *foundational* logicist project that he set himself. He failed to furnish a definition of addition that would explain the *applicability* of that operation. For the operation applies to natural numbers as *cardinal* numbers—numbers that are themselves applied as measures of size of finite collections. A prerequisite for a conceptually illuminating definition of addition on (finite) cardinal numbers is that the definition should apply also to infinite cardinals—or, at the very least, be smoothly generalizable to them (should we wish to engage in such generalization). Yet is it clear that the Dedekindian definition of addition that Frege provided cannot be generalized so as to deal with *infinite* cardinals. For, on that definition, the result of adding n to m is the result one obtains by starting with m and applying the successor operation n times. *If* (and this is a big 'if') this approach could be generalized so as to deal with Frege's number *Endlos* (or what we would today call ω) in place of m, we would obtain *ordinal* addition: for a natural number n, one would have the *ordinal sum* $\omega + n$ as the result of such adding of n to ω, rather than the *cardinal sum*, which is of course ω itself. This failure, on Frege's part, to show us how to add together *cardinal* numbers rather than *ordinal* numbers is all the more significant given his own insistence on the importance of infinite cardinals—for it is only in the infinite case that the conceptual distinction between cardinals and ordinals reveals itself in extension. In the *Grundlagen* (p. 97) Frege had praised Cantor [2] for introducing infinite numbers, writing (pp. 97–98)

> I heartily share his contempt for the view that in principle only finite Numbers ought to be admitted as actual. ... our concept of Number has from the outset covered infinite numbers as well ...

Frege was also acutely aware (loc. cit.) of the distinction between cardinal and ordinal numbers in the infinite case:

> … my terminology diverges to some extent from [Cantor's]. For my Number he uses "power", while his concept[fn] of Number [i.e. *ordinal* number—NT] has reference to arrangement in an order. Finite Numbers, certainly, emerge as independent nevertheless of sequence in series, but not so transfinite Numbers. But now in ordinary use the word "Number" and the question "how many?" have no reference to arrangement in a fixed order. CANTOR's Number gives rather the answer to the question: "the how-manyeth member in the succession is the last member?" *So that it seems to me that my terminology accords better with ordinary usage.* [Emphasis added.] If we extend the meaning of a word, we should take care that, so far as possible, no general proposition is invalidated in the process, *especially one so fundamental as that which asserts of Number its independence of sequence in series.* [Emphasis added.] For us, because our concept of Number has from the outset covered infinite numbers as well, no extension of its meaning has been necessary at all.

So Frege bequeathed to the logicist tradition a gap that by his own lights ought to be filled. It has three dimensions. To fill it, one must give a conceptually illuminating account of addition of *cardinal* numbers, which will (i) confer the correct sense on any additive numerical term; (ii) apply indifferently to both finite and infinite numbers; and (iii) afford derivations, from suitable first principles, of the Peano–Dedekind recursion equations for addition on the natural numbers. Ditto for multiplication.

It would be fatal for a genuine neo-logicism if this gap could not be filled. It is not enough to assume that the Fregean neo-logicist—let alone Frege—could rest content with having generated the natural number sequence in a satisfyingly logicist fashion, but with having thereafter finished the job simply by reprising Dedekind's account of definition (of addition and multiplication) by recursion on the natural numbers (albeit by means of definitions of addition and multiplication in terms of zero and successor). For such a move would not account for the *applicability* of the operations of addition and multiplication. Definition by recursion lives entirely on the pure mathematical side; it affords no account whatsoever of what, conceptually, addition and multiplication amount to as operations on (finite) *cardinal* numbers. The logicist wants to show what information or logical constraints concerning numbers, and the collections they number, the operations of addition and multiplication afford. Yet definition by recursion does not (by itself) explain, for example, how and why it is that if there are n Fs and m Gs, and nothing is both F and G, then the sum $n+m$ turns out to be (of necessity) the number of things that are either F or G.

Dummett ([4] p. 51), makes this point against Dedekind in what he intends as a comparison favorable to Frege. He says that Frege 'defines the sum of two *natural* numbers' [emphasis added] '… in effect, as the number of members of the union of two disjoint classes.' But Dummett provides no reference to where, exactly, Frege's alleged definition of the sum of two natural numbers is to be found.

It might be thought, on Dummett's behalf, that this 'definition' of addition is provided in *Grundegesetze* Vol. II, Section 33. This section can be ruled out, however, as a source of a satisfactory definition of addition for the logicist. The rather contorted gloss with which Frege opens his *Zerlegung* in that section shows that he is well aware that he is not really offering a definition that would meet his own adequacy criteria:

„Die Summe von zwei Anzahlen ist durch diese bestimmt“, in diesem Ausdrucke is der Gedanke des Satzes unserer Hauptüberschrift am leichtesten zu erkennen, und darum mag er angeführt sein, obwohl der bestimmte Artikel beim Subject die Aussage der Bestimmtheit eigentlich vorwegnimmt und obwohl das Wort „Summe“ hier anders gebraucht ist, als wie wir es später bei den Zahlen gebrauchen werden. Wir nennen hier nämlich [$\#x(Fx \vee Gx)$] Summe von [$\#x\,Fx$] und [$\#x\,Gx$], wenn kein Gegenstand zugleich unter den [F-] und unter den [G-] Begriff fällt.

“The sum of two numbers is determined by them”: this expression provides the easiest way to recognize the thought rendered by the formal sentence of our main heading, and for that reason this expression may be offered—despite the fact that the definite article in the subject actually anticipates the claim of determination and the fact that the word “sum” is used here differently from the way we shall later be using it in connection with the (real) numbers. That is, we call [$\#x(Fx \vee Gx)$] the sum of [$\#x\,Fx$] and [$\#x\,Gx$], if no object falls both under the concept [F] and under the concept [G].

[For Frege, *Anzahl* means cardinal number, *Zahl* means real number.]

Observe, first, that no defined symbol for addition is hereby introduced, and the ‘definition’ does not find its way into the later *Tafel der Definitionen* in the subsequent appendices. It [i.e., the *Hauptüberschrift* of Vol. II, Section 33, at p. 44], after being proved as the result labelled (‘469’ (see Vol. II, p. 58 *supra*) is listed, rather, in the *Tafel der wichtigeren Lehrsätze* (ibid., p. 274 *infra*), which follows immediately after the *Tafel der Definitionen*.

Secondly, if one were to try to turn (469 into a formal definition, one would at best obtain something along the lines (in modern notation) of

$\#x\,F_1(x) + \#x\,G_1(x) = \#x\,F_2(x) + \#x\,G_2(x)$
whenever $\#x\,F_1(x) = \#x\,F_2(x)$ and $\#x\,G_1(x) = \#x\,G_2(x)$
and nothing is both F_1 and G_1
and nothing is both F_2 and G_2.

This of course would not do as a definition of addition, because one would not be providing a sense in sufficient generality for expressions of the form $t + u$.

Thirdly, consider the second sentence of Vol. II, Section 33, ‘Wir nennen … fällt.’ It might be thought that this would furnish an explicit definition of $+$ as follows:

$$\#x\,Fx + \#x\,Gx =_{df} \#x(Fx \vee Gx) \text{ if } \neg\exists x(Fx \wedge Gx).$$

But this, too, is not a proper definition by Frege’s criteria. In Vol. II, Section 65 he writes

…das Additionszeichen ist nur erklärt, wenn die Bedeutung jeder möglichen Zeichenverbindung von der Form »$a+b$« bestimmt ist, welche Bedeutungsvolle Eigennamen man auch für »a« und für »b« einsetzen möge.

…the sign for addition is explained only when the denotation of every possible sign-combination of the form »$a + b$« is determined, no matter what meaningful proper names one might substitute for »a« and for »b«.

The 'Wir nennen' is at best a (metalinguistic) gloss concerning what one might say under rather particular circumstances (of disjointness of the concepts involved). It is not something that leads to a proper explicit definition.

Certainly, if one *had* a proper explicit definition of +, one would expect to be able to derive, as a theorem,

$$\text{If } \neg\exists x(Fx \wedge Gx), \text{ then } \#x\,Fx + \#x\,Gx = \#x(Fx \vee Gx)$$

along with a similar theorem corresponding to (469):

If $\#x\,F_1(x) = \#x\,F_2(x)$ and $\#x\,G_1(x) = \#x\,G_2(x)$
and nothing is both F_1 and G_1
and nothing is both F_2 and G_2,
then $\#x\,F_1(x) + \#x\,G_1(x) = \#x\,F_2(x) + \#x\,G_2(x)$.

We are still left empty-handed, though, in so far as we want a definition of + that would not only afford derivations from 'logical' first principles of the recursion axioms for + (and likewise for $\times$) but also accomplish the goal of providing an account of the applicability of these operations to (finite) cardinal numbers. So it does seem, on balance, as though Vol. II, Section 33 does not provide the kind of textual evidence one would like to see for Dummett's claim about addition.[2]

It would appear that *Gg.* Volume II, Section 33 is a nod from Frege in the direction of Cantor's 1895 treatment of addition and multiplication of powers in [3, Section 3], 'Die Addition und Multiplikation von Mächtigkeiten', at p. 485. Cantor contented himself with a definition of these operations that applied, as it were, *de rebus*, without providing a sense for every possible expression of the form $t + u$, where t and u are terms which, if they denote at all, denote numbers. This approach is characteristic of the mathematician who is concerned only to characterize the

---

[2] These considerations also counsel against attempting an explicit higher-order or set-theoretic definition of addition (say) that directly exploits the desired recursion equations. One might, for example, attempt to define the relation SUM(x, y, z) (which says that $z = x + y$) as the intersection of all relations $F(u, v, w)$ such that

1. $F(u, 0, u)$, and
2. If $F(u, v, w)$ and v' is the successor of v, then $F(u, v', w')$, where w' is the successor of w.

The drawback of such a proposal is that it involves either overly powerful set-theoretic commitments, or the resources of second-order logic. The definitions of addition and multiplication to be offered here, by contrast, involve the more modest resources of orderly pairing, and remain at first order. Moreover, the recursion equations, which will be derivable from the definitions, do not need to be 'built in', the way they are with the foregoing definition of SUM. Instead, the recursion equations arise as *fruitful* (albeit logically necessary) byproducts of the definitions of these operations in terms of orderly pairing, whose form is chosen with an eye to meeting the Fregean requirement of 'illumination of applicability'.

extension of an operation or of a predicate, within an abstract ontology to which intellectual access is already assumed. If one takes the numbers (or, in Cantor's case, the 'powers') as already given, then the task, understandably, is simply to give an extensionally correct characterization of the 'action' of the operation of addition on those numbers. Thus Cantor could write (loc. cit.)

> Die Vereinigung zweier Mengen M und N, die keine gemeinschaftlichen Elemente haben, wurde in Section 1, (2) mit (M, N) bezeichnet. Wir nennen sie die "*Vereinigungsmenge von M und N*".
>
> Sind M', N' zwei andere Mengen ohne gemeinschaftliche Elemente, und ist $M \sim M'$, $N \sim N'$, so sahen wir, daß auch
>
> $$(M, N) \sim (M', N').$$
>
> Daraus folgt, daß die Kardinalzahl von (M, N) nur von den Kardinalzahlen $\overline{\overline{M}} = \mathfrak{a}$ und $\overline{\overline{N}} = \mathfrak{b}$ abhängt.
>
> Dies führt zur Definition der Summe von $\mathfrak{a}$ und $\mathfrak{b}$, indem wir setzen
>
> $$\mathfrak{a} + \mathfrak{b} = (\overline{\overline{M, N}}). \qquad (1)$$

Jourdain's translation (at p. 91) is as follows:

> The union of two aggregates M and N which have no common elements, was denoted in Section 1, (2) by (M,N). We call it the "union-aggregate (*Vereinigungsmenge*) of M and N."
>
> If M′, N′ are two other aggregates without common elements, and if M∼ M′, N∼ N′, we saw that we have
>
> $$(\mathrm{M}, \mathrm{N}) \sim (\mathrm{M}', \mathrm{N}').$$
>
> Hence the cardinal number of (M,N) only depends upon the cardinal numbers $\overline{\overline{\mathrm{M}}} = \mathfrak{a}$ und $\overline{\overline{\mathrm{N}}} = \mathfrak{b}$.
>
> This leads to the definition of the sum of $\mathfrak{a}$ and $\mathfrak{b}$. We put
>
> $$(1) \qquad \mathfrak{a} + \mathfrak{b} = (\overline{\overline{\mathrm{M}, \mathrm{N}}}).$$

Certainly if one compares Frege's Vol. II, Section 33 with this Cantorian definition of addition, the affinity is unmistakeable. But whereas Cantor succeeded in the merely mathematical task of stipulating what cardinal number one would obtain by adding two cardinal numbers $\mathfrak{a}$ and $\mathfrak{b}$—a success merely *in extenso*—the Fregean logicist surely cannot rest content with that. He cannot allow himself the compromising loss of generality that is involved in having to consider *disjoint* sets M and N in the foregoing Cantorian charaterization of addition. The Fregean logicist has the more demanding (and self-imposed) task of providing a sense for *any* expression of the form $t + u$ when both t and u are terms ostensibly denoting cardinal numbers. Thus both t and u may indifferently (and independently) be numerals; parameters in natural deductions; or abstractive terms of the form $\#x F(x)$, where $F(x)$ itself may be *any* predicate of the language. And that language, note, may contain predicates in whose extensions fall physical objects. The language is not necessarily that more limited, wholly mathematical one for which Cantor plied his account of addition. Thus the Fregean logicist, in pursuit of goal (1) of applicability, would need to furnish a definition of addition that would confer the required senses upon terms such as

$$\#x(x \text{ captained England in cricket}) + \#x(x \text{ has a Philosophy degree}),$$

and

$$\#x(x \text{ was an Oxfordshire cricketer}) + \#x(x \text{ played chess for Oxfordshire}).$$

The accomplishments of Mike Brearley ensure that the former of these additive terms denotes a number at least one greater than

$$\#x(x \text{ captained England in cricket or } x \text{ has a Philosophy degree});$$

and if Paul Grice's well-known autobiographical claim is true, the latter of these additive terms denotes a number *exactly* one greater than

$$\#x(x \text{ was an Oxfordshire cricketer or } x \text{ played chess for Oxfordshire}).$$

If M is chosen to be the set of all England cricket captains and N is chosen to be the set of all Philosophy degree-holders, then Cantor's definitional method will not be able to tell us what $\overline{\overline{M}} + \overline{\overline{N}}$ is—since M and N are not disjoint. But if one takes the cardinality $\mathfrak{a}$ of the set of all England cricket captains and the cardinality $\mathfrak{b}$ of the set of all Philosophy degree-holders, then Cantor's definitional method does provide us with a route to the sum $\mathfrak{a} + \mathfrak{b}$, albeit perforce circuitously. The route is this: choose some set M with cardinality $\mathfrak{a}$ and some set N *disjoint from* M with cardinality $\mathfrak{b}$; then for $\mathfrak{a} + \mathfrak{b}$ take (in modern notation) $\overline{\overline{M \cup N}}$. Obviously in making *these* choices of M and N one will have to avoid taking for M the set of all England cricket captains and for N the set of all Philosophy degree-holders!

The circuitousness of Cantor's route to sums is neither here nor there when the sole preoccupation is that of fixing the action of the operation of addition within a domain of abstract objects (the cardinal numbers) presumed already given. Any *inter res* route from $\mathfrak{a}$ and $\mathfrak{b}$ to $\mathfrak{a} + \mathfrak{b}$ will do. But the Fregean logicist wants in addition (if one will excuse the pun) to assign a sense to *any* expression of the form $\#xF(x) + \#xG(x)$, regardless of how the extensions of the predicates $F(x)$ and $G(x)$ might fall. Each such linguistic expression must receive a sense; and in assigning such senses, one expects the sign '+' of addition to receive the same sense wherever it occurs. That, for the Fregean logicist, is the more exigent task of 'defining addition'.

The recent resurgence of interest in logicism has been based on the observation that for the purposes of a Fregean derivation of the axioms for zero and successor, and of the principle of mathematical induction, Frege's ultimately inconsistent theory of classes is not needed. It suffices to assume Hume's Principle—one of the theorems derived in Frege's system—along with other logical principles collectively consistent with it. Frege had derived Hume's Principle in the first volume of *Grundgesetze*, and all his subsequent work in pursuit of goal (2) used no more than Hume's Principle. So, a neo-logicist such as Wright contends, all that logicism needs is a defence of Hume's Principle as analytic, and the philosophical task of logicism will have been accomplished. Arithmetic will have been derived from analytic—that is to say, logical—principles.

Quite so. But that holds only in so far as the basic axioms of arithmetic really have been derived within the logicist's refurbished, and now consistent, theory, from *conceptually illuminating definitions*. The current literature on neo-logicism contains no hint as to how a logicist might define addition and multiplication in a conceptually illuminating way, and then use those definitions to provide a derivation of the recursion axioms for addition and multiplication.

Nor is there any guidance on this score from Frege himself. This is in stark contrast to his legacy of deductive trail-blazing within the easier territory of zero, successor and mathematical induction. He gave conceptually illuminating definitions, and amply suggestive deductive signpostings in the *Grundlagen*,[3] and followed them up with the captious hardscaping of the *Grundgesetze*. But that was only for zero and successor.

With addition and multiplication, matters are starkly different. The logicist searching within the two volumes of the *Grundgesetze* for any conceptual illumination of addition and multiplication will be disappointed. Frege appears not to attach any significance at all to the problem of how to characterize addition and multiplication on the whole numbers *as cardinal numbers*, and how on such a basis to justify the basic arithmetical laws governing them. Instead, he treats his reader (Vol. II, Section 65) to a remarkably unconvincing argument for the view that it would be wrong-headed to attempt to do so! He maintains that a study of arithmetical operations like addition and multiplication *restricted to the natural numbers* would sin against one of his canons of definition: that the definition of any function, such as $+$, has to provide a sense for any term of the form $a + b$, whatever *bedeutungsvolle Eigennamen* might be put in place of a and b. Thus he is setting out to accomplish the apparently more ambitious task of characterizing addition (he makes no mention of multiplication) as a function on at least the real numbers, construed as ratios of magnitudes. He gets as far as establishing the commutativity and associativity of an addition-like operation defined for his so-called *Positivalklassen*. But that still falls short of deriving even the Peano–Dedekind recursion axioms for addition alone, on the basis of a conceptually illuminating definition of addition. And besides, Frege

[3] It was these signpostings that Wright sought to make into a slightly more detailed logical itinerary in his monograph [20], which stimulated the recent revival of interest in logicism. Like Frege in the *Grundlagen*, Wright did not concern himself with addition, let alone with multiplication. But a fully detailed map of the logical terrain as it concerned only zero and successor had of course already been provided by Frege in Vol. I of *Grundgesetze*. Wright had not read the *Grundgesetze* (personal communication), because it was in German and its notation was 'rebarbative'.

himself (Vol. II, Section 157) regarded the natural numbers (*Anzahlen*) as ontologically distinct from the positive whole numbers (*positive ganze Zahlen*) that are the multiples of the unit real within a system of real numbers. So even if he succeeds in deriving laws (such as commutativity and associativity) for addition and multiplication on the latter kind of number, that would not yet vouchsafe the corresponding laws on the former kind.

This brings to a close the philosophical stage-setting for the logical and foundational contribution to follow. The logical methods to be employed centrally involve the notion of an ordered pair, taken as primitive and *sui generis*. It is by means of the logical notion of orderly pairing that the logicist's gap can be filled.

2 Orderly Pairing

An ordered pair is a 'thing of the form' $\langle t, u\rangle$, in which the order of the things paired *matters*. It is this ordering which first gives us the first 'member', and then gives us the second 'member'. So the ordered pair $\langle t, u\rangle$ has t as its first 'member' and u as its second 'member'. The word 'member' is used in scare quotes in order to flag the fact that it is not to be assumed that we are dealing with membership in the sense of set theory. The ordered pair $\langle t, u\rangle$ will be identical to the ordered pair $\langle u, t\rangle$ if, but only if, $t = u$.

In order to avoid any possible future confusion between talk of members of ordered pairs and talk of set-theoretic members of *sets*, reference henceforth will be to the (first and second) *coordinates* of an ordered pair.

We are used to thinking of ordered pairs within the confines of modern set theory. The definition we use today, due to Kuratowski [12] at pp. 170–171, is

$$\langle t, u\rangle =_{df} \{\{t\}, \{t, u\}\}.$$

Against the background of set theory, the Kuratowski definition of ordered pair meets the following essential condition of adequacy: it allows unambiguous specification of both the *first* and the *second* coordinate of any given ordered pair. With the Kuratowski pair-set, which is the set-theoretic surrogate for the ordered pair, we can define the ordered pair's first coordinate as

> that thing that is an element of each member of the pair-set.

The ordered pair's second coordinate can then be defined as follows:

> if the pair-set's union has just one member, then that member; otherwise, that member of the pair-set's union that is not the first coordinate.

By contrast with Kuratowski's set-theoretic reduction of ordered pairs, this study is concerned to deal directly with orderly pairing as an operation *sui generis*. The values of that operation, the ordered pairs themselves, are likewise individuals *sui generis*.

Given the history of set theory and of proof theory, there was never an appropriate juncture at which it would have been natural to pursue the question of introduction and elimination rules for both orderly pairing and the projection of coordinates. For by the time Gentzen first produced his systems of natural deduction in 1936, set theorists had already settled on Kuratowski's definition of ordered pair, on von Neumann's definition of the ordinals, and on the definition of arbitrary functions as sets of ordered pairs. These very serviceable definitions fulfilled all the modelling demands of mathematics, from a structuralist point of view. So no one gave much thought to the question whether certain notions that had been furnished with serviceable set-theoretical surrogates might not be better characterizable directly, as *sui generis*, and as obeying a genuine *logic* of their own—a logic best laid out as a system of Gentzen-style introduction and elimination rules for the notions concerned. Sections (3–10) aim to do that for the orderly pairing operation and its associated coordinate-projection operations. Sections (11–14) will then put orderly pairing to work in our search for conceptually illuminating primitive rules governing addition and multiplication of cardinal numbers. These rules serve to define those two notions, and afford natural proofs of the Dedekind–Peano recursion axioms. (The task of laying out those proofs, however, will have to be deferred to another occasion.)

3 Notation

At this point let us give up the usual notation

$$\langle t, u\rangle$$

for an ordered pair, and adopt instead the notation

$$\pi(t, u).$$

The new notation stresses that the formation of an ordered pair is a matter of applying a binary function (of 'orderly $\underline{p}$*airing*') to two arguments in a given order. The value of the function will, in general, depend both on the arguments and on the order in which they are given. The value will be an *individual* in the domain of discourse. (Remember that we are dealing here with a proposed extension of first-order logic, by focusing on certain functions.) One might be tempted to regard this individual $\pi(t, u)$ as somehow *composed out of* the arguments t and u. But that would be a mistake, in so far as we are concerned here with how to understand the first-order *logic* of the orderly pairing function π. All that one needs to grasp is that the function π maps t and u to their ordered pair. Exactly what kind of entity this ordered pair might be is a question that need not be pressed for the time being. It is enough to know that the ordered pair is an individual in the domain, falling within the range of the quantifiers.

A universally free logic will be used in all that follows, as in [16]. This logic is based on the Russellian construal of the truth-conditions of atomic predications, including identities: for an atomic predication $A(t_1, \ldots, t_n)$ to be true, all the terms $t_1, \ldots, t_n$ must denote (and of course the relation A must obtain among their denotations). One can express the fact that the term t denotes by writing $\exists x(x = t)$. One can abbreviate this as $\exists!t$. So we have the so-called

Rule of Denotation for Atomic Predications

$$\frac{A(t_1, \ldots, t_n)}{\exists!t_i} \quad (1 \leq i \leq n)$$

There is also a rule designed to capture the idea that in order for a functional term to denote, its argument terms must denote.

Rule of Denotation for Functional Terms

$$\frac{\exists!f(t_1, \ldots, t_n)}{\exists!t_i} \quad (1 \leq i \leq n)$$

Note that since we must allow in general for *partial* functions, i.e. functions that are not everywhere defined, our free logic does *not* have the rule

$$\frac{\exists!t_1 \quad \ldots \quad \exists!t_n}{\exists!f(t_1, \ldots, t_n)}$$

But the orderly pairing function π is special: it *is* total. This is expressed by the so-called Rule of Totality (for π) below. The Rule of Totality tells one that from *any* two individuals, in *any* order, one can form the corresponding ordered pair. Of course, the expression ‘one can form’ is merely metaphorical. We do not compose ordered pairs out of their coordinates. Rather, ordered pairs exist as abstract objects, given only that their coordinates exist. That is to say, any two things can be paired, in any of their possible orders.

In addition to the pairing function π, one needs to treat the *coordinate projection functions* λ (for the first, or *left* coordinate) and ρ (for the second, or *right* coordinate). These are monadic functions. They are defined only on ordered pairs. So they are not necessarily total.

4 The Formal Rules for Orderly Pairing

For the purposes of providing introduction and elimination rules, we focus on the projection functions λ (projecting the left coordinate) and ρ (projecting the right coordinate). The introduction rule for λ tells us the canonical conditions under which one can infer an identity of the form $t = \lambda(u)$; likewise (*mutatis mutandis*) with the introduction rule for ρ.

$$(\lambda\text{-}I) \quad \frac{u = \pi(t, v)}{t = \lambda(u)}$$

$$(\rho\text{-}I) \quad \frac{u = \pi(v, t)}{t = \rho(u)}$$

Corresponding to these introduction rules for the projection operations λ and ρ are their elimination rules.[4]

$$(\lambda\text{-}E) \qquad \frac{t = \lambda(u) \qquad \begin{array}{c} \overline{u = \pi(t, a)}^{(i)} \\ \vdots \; a \text{ parametric} \\ \psi \end{array}}{\psi}(i)$$

$$(\rho\text{-}E) \qquad \frac{t = \rho(u) \qquad \begin{array}{c} \overline{u = \pi(a, t)}^{(i)} \\ \vdots \; a \text{ parametric} \\ \psi \end{array}}{\psi}(i)$$

These elimination rules match their respective introduction rules, in that they carry existential import. Consider, for example, the rules for λ. The introduction rule for λ tells us that the only way to prove $t = \lambda(u)$ is to prove that u is an ordered pair whose left coordinate is t. But that requires that there be some right coordinate (called v in the introduction rule) that, with the left coordinate t, makes up the ordered pair u. The elimination rule for λ then exploits this existential assumption concerning the right coordinate. (*Mutatis mutandis* for ρ and the left coordinate.) The harmony between the introduction and elimination rules for λ is brought out by the following very obvious reduction procedure. It gets rid of any sentence oc-

[4] The reader unfamiliar with proof-theoretic notation is advised that the parenthetically enclosed numeral '(i)' has an occurrence labelling the step at which the indicated assumption-occurrences higher up at 'leaf nodes' of the sub-proof(s) are *discharged* by applying the rule in question. A discharged assumption no longer counts among the assumptions on which the conclusion of the newly created proof depends. Also, we say that a is *parametric* within a sub-proof just in case among the undischarged assumptions, and conclusion, of the sub-proof in question, a occurs only within sentences of the indicated form.

currence that stands as the conclusion of a step of λ-introduction and as the major premise of λ-elimination:

$$\dfrac{\dfrac{\begin{array}{c}\Sigma\\ u=\pi(t,v)\end{array}}{t=\lambda(u)} \qquad \begin{array}{c}\overline{u=\pi(t,a)}^{\,(i)}\\ \Theta\\ \psi\end{array}}{\psi}\,(i) \quad \longmapsto \quad \begin{array}{c}\Sigma\\ (u=\pi(t,v))\\ \Theta[a/v]\\ \psi\end{array}$$

The operation of orderly pairing can be applied to *any* two things that exist. That is to say, the function π is everywhere defined (or total):

$$\textbf{Rule of Totality} \qquad \frac{\exists!t \quad \exists!u}{\exists!\pi(t,u)}$$

5 Derivable Results

The formal rules of the previous section suffice for proofs of the following central results. Proofs are given in the Appendix.

1. $\dfrac{u=\lambda(t) \qquad v=\rho(t)}{t=\pi(u,v)}$
2. $\dfrac{u=\lambda(t)}{t=\pi(u,\rho(t))} \quad \dfrac{v=\rho(t)}{t=\pi(\lambda(t),v)}$
3. $\dfrac{\exists!\pi(t,u)}{\exists!t} \quad \dfrac{\exists!\pi(t,u)}{\exists!u}$
4. $\dfrac{\pi(t,u)=\pi(v,w)}{t=v} \quad \dfrac{\pi(t,u)=\pi(v,w)}{u=w}$
5. $\dfrac{t=v \qquad u=w}{\pi(t,u)=\pi(v,w)}$
6. $\dfrac{\exists x\ \pi(x,t)=u}{t=\rho(u)} \quad \dfrac{\exists x\ \pi(t,x)=u}{t=\lambda(u)}$
7. $\dfrac{t=\rho(u)}{\exists x\ \pi(x,t)=u} \quad \dfrac{t=\lambda(u)}{\exists x\ \pi(t,x)=u}$
8. $\dfrac{\exists!\lambda(t)}{\exists!\rho(t)} \quad \dfrac{\exists!\rho(t)}{\exists!\lambda(t)}$

Result (8) is called the Principle of Parity.

Results (4) and (5) can be combined into an abstraction principle that Kanamori [11] calls *Peano Pairing*:

$$\pi(t,u)=\pi(v,w) \leftrightarrow (t=v \wedge u=w).$$

6 Remarks on the Rules

6.1 Rules of Natural Deduction in the Style of Gentzen and Prawitz

The only functions for which we have provided explicit introduction and elimination rules (in the style of Gentzen [8] and Prawitz [14])[5] are the two unary projection functions λ and ρ. Note, however, that the binary orderly-pairing function π does indeed have its own introduction and elimination rules. The introduction rule for π is

$$(\pi\text{-}I) \quad \frac{u = \lambda(t) \quad v = \rho(t)}{t = \pi(u, v)},$$

which is our Result (1) above. The elimination rules for π are

$$(\pi\text{-}E) \quad \frac{t = \pi(u, v)}{v = \rho(t)} \quad \frac{t = \pi(u, v)}{u = \lambda(t)}$$

which are none other than the rules (ρ-I) and (λ-I) respectively. The reduction procedure for these π-rules is obvious.

It is not possible to state adequate rules for the projection operations λ and ρ that do not involve mention of the operation π of orderly pairing. Nor is it possible to state adequate rules for π that do not mention λ and ρ. These three operations are conceptually *coeval*. It is therefore impossible to force one's introduction and elimination rules into a form where the operator concerned is the *sole* operator occurring in the statement of the rule. This, however, should not count against the rules that we have given as capturing the *logical* content of these operations.

The introduction and elimination rules for λ and ρ cannot be derived from the introduction and elimination rules for π alone. They can, however, be derived from those for π *together with the Principle of Parity*. Here, for example, is the derivation of (λ-E) using Parity in the direction from $\exists!\lambda(u)$ to $\exists!\rho(u)$:

[5] Gentzen and Prawitz did not, themselves, consider the matter of natural deduction rules governing *term*-forming operators. They dealt only with *sentence*-forming logical operators such as connectives and quantifiers. The current treatment of introduction and elimination rules for term-forming operators is in the style introduced in [16], and further developed in [17–19]. The leading idea is to frame an introduction rule for a term-forming operator Ω by specifying the conditions under which one would be justified in asserting an identity of the general form '$t = \Omega \ldots$', where we focus on one dominant occurrence of Ω (here, on the right-hand side). Likewise, one frames the corresponding elimination rule by using such an identity as the major premise, and specifying what one would be warranted in inferring from its assertion (presumed justified). A full statement of the rules for constructive logicism is to be found in Section 14 below.

$$
\begin{prooftree}
\AxiomC{$t=\lambda(u)$}
\AxiomC{$t=\lambda(u)$}
\UnaryInfC{$\exists!\lambda(u)$}
\UnaryInfC{$\exists!\rho(u)$}
\UnaryInfC{$\rho(u)=\rho(u)$}
\BinaryInfC{$u=\pi(t,\rho(u))$}
\AxiomC{$t=\lambda(u)$}
\UnaryInfC{$\exists!\lambda(u)$}
\UnaryInfC{$\exists!\rho(u)$}
\BinaryInfC{$\exists x\ u=\pi(t,x)$}
\AxiomC{}
\RightLabel{(i)}
\UnaryInfC{$u=\pi(t,a)$}
\noLine
\UnaryInfC{$\vdots$}
\noLine
\UnaryInfC{ψ}
\RightLabel{(i)}
\BinaryInfC{ψ}
\end{prooftree}
$$

In similar fashion, one can derive (ρ-E) using Parity in the direction from $\exists!\rho(u)$ to $\exists!\lambda(u)$.

As we have also seen, the introduction and elimination rules for π, and the Principle of Parity, can be derived from the introduction and elimination rules for λ and ρ. The latter introduction and elimination rules, therefore, are a simpler primitive set to adopt. But we *could*, if we wished, adopt as primitive instead the introduction and elimination rules for π, along with the Principle of Parity.

Regardless of which of these two alternatives we adopted, we would still need to postulate the Principle of Totality, which is used for the proof of Result (5), the right-to-left direction of Peano Pairing.

6.2 Fregean Abstraction

The logic of orderly pairing could be cast in a more traditional Fregean form, rather than in the form of introduction and elimination rules in the style of Gentzen and Prawitz. Fregean abstraction (for a term-forming operator Ω) involves laying down necessary and sufficient conditions, not mentioning Ω, for the truth of an identity statement involving *two* Ω-terms rather than one. In the case at hand, where we take π for Ω, such a statement of necessary and sufficient conditions is Peano Pairing itself:

$$\text{(PP)} \qquad \pi(t,u)=\pi(s,v) \quad \leftrightarrow \quad (t=s \wedge u=v).$$

The right-hand side of this biconditional involves no mention of π.

This quantifier-free statement of Peano Pairing as a Fregean abstraction principle is implicitly universally quantified:

$$\forall x \forall y \forall z \forall w\ [\ \pi(x,z)=\pi(y,w) \quad \leftrightarrow \quad (x=y \wedge z=w)\ ].$$

Taking the direction from right to left, we obtain an easy proof of the Principle of Totality:

$$\dfrac{\dfrac{\dfrac{\dfrac{\exists!a}{a=a} \qquad \dfrac{\exists!b}{b=b}}{a=a \wedge b=b} \qquad \dfrac{\exists!b \quad \dfrac{\exists!b \quad \dfrac{\exists!a \quad \dfrac{\exists!a \quad \forall x \forall y \forall z \forall w[\pi(x,z)=\pi(y,w) \leftrightarrow (x=y \wedge z=w)]}{\forall y \forall z \forall w[\pi(a,z)=\pi(y,w) \leftrightarrow (a=y \wedge z=w)]}}{\forall z \forall w[\pi(a,z)=\pi(a,w) \leftrightarrow (a=a \wedge z=w)]}}{\forall w[\pi(a,b)=\pi(a,w) \leftrightarrow (a=a \wedge b=w)]}}{[\pi(a,b)=\pi(a,b) \leftrightarrow (a=a \wedge b=b)]}}{\pi(a,b)=\pi(a,b)}}{\exists!\pi(a,b)}$$

We can also *define* the two projection functions λ and ρ in terms of the pairing function π, by using the definite description operator ι:

$$\begin{aligned} \lambda(t) \quad &=_{df} \ \iota x \exists y \ t=\pi(x,y); \\ \rho(t) \quad &=_{df} \ \iota y \exists x \ t=\pi(x,y). \end{aligned}$$

Using these definitions, we can derive the introduction and elimination rules for λ and ρ. Note that if t is not an ordered pair, then both $\lambda(t)$ and $\rho(t)$ will be non-denoting terms. And that is how it should be.

It is interesting that orderly pairing is one of the cases in which the method of Fregean abstraction actually works, without incurring paradox (or well-grounded worries about potential paradox). This is because of intuition (i) below. Fregean abstraction applied to the notion of *set*, as opposed to orderly pairing, incurs paradox. Here the Fregean abstraction principle would be

$$\{x \mid Fx\} = \{x \mid Gx\} \ \leftrightarrow \ \forall x(Fx \leftrightarrow Gx).$$

This would have the special case

$$\{x \mid Fx\} = \{x \mid Fx\} \ \leftrightarrow \ \forall x(Fx \leftrightarrow Fx).$$

Since the right-hand side is a theorem, we can detach and conclude

$$\{x \mid Fx\} = \{x \mid Fx\},$$

hence

$$\exists!\{x \mid Fx\},$$

regardless of what the predicate Fx might be. Taking $\neg(x \in x)$ for Fx, we obtain Russell's Paradox.

7 Analytic Intuitions

The *logic* of orderly pairing (and projection of coordinates) is captured by any sufficiently strong subcollection of the foregoing rules. That is, the rules in question capture what is essential to the *concept* of orderly pairing. The following five ingredients are essential, and, collectively, they exhaust the concept in question:

(i) any two things can be paired, and in either order;
(ii) not necessarily everything is an ordered pair;
(iii) two ordered pairs are identical just in case their respective coordinates are;
(iv) an ordered pair exists only if both its coordinates do; and
(v) a thing can have the one kind of coordinate only if it has the other kind as well (that is, the projection functions are defined only on ordered pairs, and indeed on exactly the same ones, namely all of them).

These intuitions (i)–(v) are captured by any of the three rule-collections that we have considered above.

But there is another source of intuitions about orderly pairing that needs to be addressed. These are intuitions about the global properties of the operation of orderly pairing. They are addressed in the next section.

The situation at which we have thus far arrived in formulating a logic for orderly pairing corresponds to the situation one is in with what I have called the 'logic of sets', after formulating introduction and elimination rules for the set-abstraction operator. (See Tennant [16], Chapter 7, and [19].) Those rules capture the analytic connections among set-abstraction, membership and possession of defining properties (of sets). In particular, they reveal the axiom of extensionality as a *consequence* of a deeper analysis of the notion of set, as captured by the introduction and elimination rules. The rules are mute, however, on the question of what sets actually exist, either outright or conditionally. This is because the rules make up a *free* logic. The introduction and elimination rules for the set-abstraction operator provide a conceptual analysis of the notion of set, but not an ontological theory about what sets actually exist. The ontological questions are the ones properly left to set *theory*, as opposed to the logic of sets (or what Quine called 'virtual set theory'). Indeed, one might venture a little further than the logic of sets, by laying down rules (or axioms) for the *hereditarily finite sets*—of which, of course, there are infinitely many. One could take the view that the existence of the empty set is an analytic matter, as would be also the conditional existence of pair sets, unions, and power sets. But it would not be analytic that any particular infinite set exists.

This extension of the 'analytic logic of sets' would be a closer analogue to what we have been doing with the logic of orderly pairing. For, in the latter, we have commited ourselves to the existence of the ordered pair of any two existents. Provided, then, that there are at least two existents, it follows right away that there will be infinitely many. (This is so even if we leave the so-called *problem of conflation* unresolved—for which, see below.)

There is an interesting difference, however, between the logic of orderly pairing and the logic of sets, when it comes to considering the use that we make of them, respectively, for mathematics. As far as the notion of ordered pair is concerned, all that is important is already captured in the *logic* of orderly pairing. We do not need to settle the further 'ontological' questions addressed in the next section before having a serviceable notion of ordered pair.

8 Ontological Intuitions

According to Kanamori ([11], p. 289),

> (κ) Peano Pairing is 'the instrumental property which is all that is required of the ordered pair.'

The word 'instrumental' here adverts to the needs of mathematics. Mathematics needs ordered pairs that behave in a certain minimally constrained way, a way that will allow for a definition of functions as sets of ordered pairs, etc. Once that minimally constrained behavior is guaranteed (by combination, say, of set-theoretical axioms and a definition of ordered pairs as Kuratowskian pair-sets), the mathematician's needs are satisfied. It is then superfluous, as far as mathematics is concerned, to inquire further into the nature of ordered pairs themselves, *were they to be taken as sui generis.* Now, as it happens, set theory contains the Axiom of Foundation, which ensures that the set-membership relation $\in$ is well-founded. Set-theoretically defined ordered pairs therefore acquire a well-founded pedigree of ordered-pair formation from the well-foundedness of $\in$. But this does not provide a principled answer to the question whether pedigrees of ordered-pair formation should be well-founded when the operation of orderly pairing is taken as *sui generis*, and not as set-theoretically defined.

This problem is especially acute when we consider that the Rule of Totality seems so natural. Any two things can be paired in orderly fashion. The resulting ordered pair will be distinct from each of them; and will in turn be eligible to be a coordinate of yet other ordered pairs. That there is such an operation, in thought at least, seems undeniable. Note, however, that it would be impossible to render this ambitious thought within the confines of set theory. For then one would have to construe the operation of orderly pairing as a function, mapping objects a and b (in that order) to the ordered pair $\langle a, b\rangle$. The function would be construed as a set of ordered pairs whose first coordinate represents the input, and whose second coordinate represents the output. The input to any binary function is actually an ordered pair; and it would be the first coordinate of an ordered pair ('in' the graph of the function) whose second member would be the output of the function, i.e. the ordered pair itself. Hence the orderly pairing function would end up being the set of all ordered pairs of the form $\langle\langle a, b\rangle, \langle a, b\rangle\rangle$, where a and b are arbitrary individuals in the domain (of set theory). But no such set can exist, at least not according to a theory such as ZFC. Rather, it would have to be construed as a proper class. Even the theory NBG, however, would not throw much light on the operation of orderly pairing, construed now as the *proper class* of all ordered pairs of the form $\langle\langle a, b\rangle, \langle a, b\rangle\rangle$, where a and b are arbitrary individuals in the domain. For, the Rule of Totality, as it governs the operation of orderly pairing *sui generis*, would apply to any two *proper classes*. If the latter are *things* in the domain (which, according to NBG, they are), then, it seems, we can pair them in whichever order we please. It would seem, then, that this operation of orderly pairing cannot even be captured as a *proper class* of ordered pairs.

What we are seeing here is a special case—involving the operation of orderly pairing—of a more general problem that afflicts the set-theoretic aspiration to provide a set-theoretic surrogate (i.e., a *set*) for any mathematical entity, including operations that are everywhere defined. Within set theory itself, for example, we have the operation of power-set formation, which is everywhere defined. (This is so even when we allow for *Urelemente*. The power set of any *Urelement* is the singleton of the empty set.) The power-set operation $\wp$ cannot *be* a set (of ordered pairs), since it would be too large. Yet we can use the functional expression $\wp(x)$ in set theory without risk of contradiction or paradox. This is because every occurrence of a term of the form $\wp(t)$ can be replaced by a set-abstract of the form

$$\{x \mid \forall y(y \in x \rightarrow y \in t)\}.$$

Because power-set formation is thus explicitly definable in set-theoretic terms, we are never at any theoretical 'loss for words' when talking about power sets, even though, on pain of contradiction, we do suffer an *ontological* deficit in not being able to identify any particular *set* within the set-theoretic universe as 'being' the operation of forming power sets. We can reason set-theoretically 'as if' $\wp$ were a primitive operation, *sui generis* and everywhere defined, because we know that anything we might wish to say in terms of $\wp$ can be translated, via the foregoing definition, into consistent set-theoretical terms. Having the lexical primitive $\wp$ affords a mere *conservative extension* of ordinary set theory, i.e. of first-order set theory formulated only in terms of set (abstraction and) membership. What we have just observed concerning the operation of power-set formation applies also to the operations of union and of pair-set formation. These operation are term-wise definable within set theory. So having primitive functional symbols to express them would result in only a conservative extension of ordinary set theory.

Let us return now to ordered pairing. Kanamori's claim (κ) may well be allowed by one who insists only on respecting the minimally demanding structural intuitions about ordered pairs in mathematical contexts where ordered pairs are simply devices for constructing set-theoretical surrogates of mathematical objects of various more complicated kinds. But the claim may be questioned by one who wishes to know more about the essential natures of ordered pairs, when orderly pairing is taken as *sui generis*.

Two questions can be asked in this regard:

(i) If one decomposes an ordered pair by projecting its coordinates, and repeats this operation on coordinates that are themselves ordered pairs, could one continue *ad infinitum*?—or is the 'pedigree' of any ordered pair finite?

(ii) Can such a ‘pedigree’, even if finite, nevertheless contain loops? That is, could a finite sequence of applications of λ and/or ρ to any ordered pair t produce t as its result?

The answer to (i) should be that the ‘pedigree’ of any ordered pair should indeed be finite. This is an *ontological* intuition about ‘grounding’. The thought is that the operation of orderly pairing cannot be indefinitely iterated; it must have *begun* with things that are *not* ordered pairs. (Let us call such things *monads*, with no echo of Leibniz. Monads might be *Urelemente*, but they could also be sets. The set ω, for example, is a monad, since it is not an ordered pair.) Hence also the answer to (ii) is that there cannot be any loops in the pedigree of an ordered pair. That pedigree must be a finite *tree*.

These ontological intuitions go strictly beyond what is required of the notion of ordered pair in order for it to serve as it does in the set-theoretical reconstruction of mathematics. Given the Axiom of Foundation in set theory, and the Kuratowski definition of ordered pair, it is clear that the ‘orderly pairing’ pedigrees of Kuratowskian ordered pairs must be finite trees. But note that this is only because the membership relation among *sets* is in general well-founded. If one is dealing with orderly pairing *sui generis*, outside the context of set theory, then one has to do more in order to ensure that the pedigrees of one’s ordered pairs are indeed finite trees. This can be seen from the examples given in the next section.

9 Deviant Interpretations

Surprisingly, the foregoing rules for the orderly pairing operation π, and its associated projection functions λ and ρ, underdetermine the notion of ordered pair to the extent that it is possible to identify certain ordered pairs with the individuals from which they are formed. Suppose given two distinct individuals a and b. Our rules force the four ordered pairs $\langle a, a\rangle$, $\langle a, b\rangle$, $\langle b, a\rangle$, and $\langle b, b\rangle$, to be distinct from each other. Those rules do not, however, force all four of these ordered pairs to be distinct from a and to be distinct from b. The following arrangement, for example, is compatible with our rules:

$$\begin{array}{cccc} \langle a, a\rangle & \langle a, b\rangle & \langle b, a\rangle & \langle b, b\rangle \\ \bullet & \bullet & \bullet & \bullet \\ a & b & & \end{array}$$

The reader can easily check that the individual a could be held identical to any one of the four ordered pairs that can be formed from a and b, and the individual b could be held identical to any one of the remaining three. The resulting denotation diagram is easily extendible to a model for our rules. (Extension of course is required in order to cope with the indefinitely many further possible iterations of orderly pairing.) Let us call this *the problem of conflation*.

Observe also that the one-element model satisfies our rules. In this model, consisting of a single individual a, say, all ordered pairs are identical to a. These include $\langle a, a\rangle$, $\langle a, \langle a, a\rangle\rangle$, $\langle\langle a, a\rangle, a\rangle$, $\langle\langle a, a\rangle, \langle a, a\rangle\rangle$, etc. This is a very degenerate and extreme case of the problem of conflation.

Even if the problem of conflation were to be solved by further stipulation of (axioms or) rules, another problem would remain: there could be infinitely descending paths of successive projections of coordinates. Let us call this the *problem of regress*.

Here is an example. Think of an infinite binary tree with root a. The nodes in this tree will be ordered pairs. Each node's daugher nodes will be its left and right coordinates. The operation of taking a left or a right coordinate can be iterated indefinitely. This seems strongly counterintuitive to anyone who thinks that ordered pairs must in some sense ultimately be 'built up out of' things that are not, themselves, ordered pairs. The claim 'Everything is an ordered pair' is consistent with our rules thus far, and would indeed force all interpretations to be deviant. But even the denial of that claim would not suffice to eliminate the problem of regress. For let b be something that is not an ordered pair. Add b to the domain consisting of the earlier infinite binary tree rooted on a. Then 'close' the domain with all the further ordered pairs that can be generated from the individuals already at hand. The result is a model of all our rules thus far, plus the axiom 'Not everything is an ordered pair.' And since this model contains the infinitely descending branches of the infinite binary tree rooted on a, the problem of regress has not gone away. Indeed, just one such branch would pose the problem.

So, if one wishes to respect the ontological intuitions expressed above, one will have to lay down more rules.

10 Supplementing the Rules

The simplest supplemental rules will rule out the one-element model for the logic of orderly pairing:

$$\frac{t = \lambda(t)}{\bot} \qquad \frac{t = \rho(t)}{\bot}$$

So these rules would solve the extreme form of the problem of conflation. But they would not be enough to solve the problem of conflation in general. For, while it would ensure that a and $\langle a, a\rangle$ are distinct, one would still be able to construct models in which other undesirable conflations occur. To be sure, the distinct individuals a and b would now each have to be distinct from each of the four ordered pairs that can be formed from them. These six individuals would give rise to as many as thirty further ordered pairs. But nothing in the rules thus far can prevent a from being held identical to, say, $\langle\langle a, a\rangle, b\rangle$ and b from being held identical to, say, $\langle a, \langle b, b\rangle\rangle$, and letting the other twenty eight ordered pairs be new distinct individuals. In this (ontologically) deviant interpretation, the left (and right) coordinate of the left coordinate of a would be a itself; and the right (and left) coordinate of the right coordinate of b would be b itself. That is, there would be two loops (of length 2) in the orderly-pairing pedigree of a; and likewise for b.

These ontologically deviant interpretations could be ruled out by prohibitions more powerful than the ones just given, to the effect that pairs cannot be either of their coordinates. That prohibition merely rules out shortest possible loops, of length 1, in orderly-pairing pedigrees. What is needed now is a way of ruling out all possible (finite) loops.

This could be done by adopting the family of rules of the following form (for $n \geq 1$):

$$\frac{\gamma_1(t_1) = t_2 \quad \gamma_2(t_2) = t_3 \quad \ldots \quad \gamma_n(t_n) = t_1}{\bot},$$

where each γ_i is either λ or ρ, and $n \geq 1$.

An alternative, and equivalent, set of schemes would be

$$\frac{\gamma_n(\gamma_{n-1}(\ldots \gamma_1(t) \ldots)) = t}{\bot}$$

This would solve the problem of conflation. But the problem of regress still remains. Perhaps the best way to solve the problem of regress is to lay down a principle of induction. Recall our earlier notion of a *monad*, i.e. something that is not an ordered pair. The following are definitional rules of reflection for that notion:

$$(M\text{-I}) \quad \begin{array}{c} \overline{t = \pi(a, b)}^{(i)} \\ \vdots \\ \underline{\quad \bot \quad}_{(i)} \\ M(t) \end{array} \quad a, b \text{ parametric} \qquad\qquad (M\text{-E}) \ \frac{t = \pi(u, v) \qquad M(t)}{\bot}$$

Equipped with the notion of a monad, we can now state a principle of induction for orderly pairing, in which a and b are parametric within the subordinate proofs in which they occur:

$$\frac{\begin{array}{c} \overline{M(a)}^{(i)} \\ \vdots \\ \Phi(a) \end{array} \qquad t = \pi(u, v) \qquad \begin{array}{c} \underbrace{\overline{\Phi(a)}^{(i)} \quad \overline{\Phi(b)}^{(i)} \quad \overline{\exists! a}^{(i)} \quad \overline{\exists! b}^{(i)}} \\ \vdots \\ \Phi(\pi(a, b)) \end{array}}{\Phi(t)}\,(i)$$

Here, $\Phi(x)$ is any first-order formula with x as its sole free variable. So the Principle of Induction is a first-order rule-schema.

11 Applications of the Logic of Orderly Pairing in a Constructive Logicist Account of Arithmetic

The constructive logicist account of number in [17] was furnished by means of a formal language of first-order logic with the variable-binding term-forming operator $\#x\Phi(x)$ ('the number of Φs'), the name 0 and the unary function sign s ('successor'). Natural deduction rules were laid down for these expressions, in a way that arguably captures their intended meanings. (See Section 14 below for a statement of all those rules in their current preferred form.) Detailed formal derivations were given for the Peano–Dedekind postulates governing 0 and s, and these derivation were carried out in intuitionistic relevant logic. This constructive logicist account also delivers every instance of *Schema N*:

$$\#x F(x) = \underline{n} \leftrightarrow \text{ there are exactly } n \ F\text{s},$$

thereby meeting what was proposed as a material adequacy condition on any theory of the natural numbers. In Schema N, $\underline{n}$ is the numeral for the number n, and the right-hand side contains no reference to numbers. An illustrative instance would be

$$\#x F(x) = ss0 \leftrightarrow \exists x \exists y (\neg x = y \wedge \forall z (Fz \leftrightarrow (z = x \vee z = y))).$$

The material adequacy condition ensures, in Fregean spirit, that the account captures the general *applicability* of natural numbers (for counting finite collections of things, concrete or abstract). The derivation of the Peano–Dedekind postulates ensures that the account captures the pure mathematics of the natural numbers (as abstract entities considered in their own right).

The account did not extend, however, to addition and multiplication of natural numbers. An appropriate extension is to be offered here. One can exploit the resources of the logic of orderly pairing in order to furnish adequate rules governing both addition and multiplication. The aim is to provide a rule-theoretic analysis that is conceptually lucid and convincing, and that affords straightforward deductions of the Peano–Dedekind postulates governing addition and multiplication. (Here it should be borne in mind that [17] has already secured the Peano–Dedekind postulates governing zero and successor, in a system based on similar rules.)

It is important to appreciate at the outset that the aim is not to furnish pair-theoretic surrogates for the natural numbers, in the way that set theory furnishes set-theoretic surrogates for them (namely, the finite von Neumann ordinals). The natural numbers will remain as objects of abstraction, *sui generis*. In the terminology of the present study, numbers will be monads, not ordered pairs. So we resist the temptation to exploit the operation of orderly pairing in order to furnish a surrogate for the successor function, such as

$$s(n) =_{df} \pi(0, n).$$

On such an approach, as soon as one had secured the existence of zero, the succeeding 'numbers' 1, 2, 3, … would be generated as

$$\pi(0, 0),\ \pi(0, \pi(0, 0)),\ \pi(0, \pi(0, \pi(0, 0))),\ \ldots$$

But such a definitional fix is to be eschewed. The numbers have, after all, already been obtained in their own right (cf. [17]), as the necessarily existing *Bedeutungen* of arithmetical terms (both pure and applied) that obey appropriate sense-constituting rules of inference. That much is achieved simply in securing the Peano–Dedekind postulates governing zero and successor. Ordered pairs will, however, feature in the current *extension* of what was accomplished in [17], an extension intended to furnish a logicist analysis of the operations of addition and multiplication. In providing the sought extension, we can avail ourselves of the *logic* of orderly pairing. The extended account will still be *logicist*, both in spirit and in letter. Details will emerge below.

11.1 A First (but Inadequate) Stab at Rules for Addition

Here is a first and obvious way to try to characterize addition, using only the resources of the constructive logicist account of [17].

If the number of Fs is t, and the number of Gs is u, and nothing is both F and G, then the number of things that are either F or G is the *sum* $(t + u)$. This simple-minded thought is captured by the following rule:

$$\dfrac{\#xFx = t \qquad \#xGx = u \qquad \begin{array}{c} \underbrace{{}^{(i)}\overline{Fa}\ ,\ \overline{Ga}^{(i)}} \\ \vdots \\ \bot \end{array} \quad a \text{ parametric}}{\#x(Fx \vee Gx) = (t + u)}\,(i)$$

As a special case, we would be able to derive the result that if the number of Fs and the number of Gs exist, and nothing is both F and G, then the sum of those numbers is the number of things that are F *or* G:

$$\dfrac{\exists!\#xFx \qquad \exists!\#xGx \qquad \begin{array}{c} \underbrace{{}^{(i)}\overline{Fa}\ ,\ \overline{Ga}^{(i)}} \\ \vdots \\ \bot \end{array} \quad a \text{ parametric}}{\#x(Fx \vee Gx) = (\#xFx + \#xGx)}\,(i)$$

This way of characterizing addition, however, suffers from the limitation that F and G must be disjoint concepts. Yet even for overlapping concepts F and G one should be able to make sense of the sum of $\#xFx$ and $\#xGx$. So the characterization of addition afforded by the foregoing rule seems insufficiently general.

It is insufficiently general on another score too: how is it to be used so as to capture $t + u$ when both t and u are (pure) numerals? The most obvious way to furnish a concept F with exactly t satisfiers is to take Fx to be 'x is a natural number preceding t'. (Likewise, take Gx to be 'x is a natural number preceding u'.) But that brings with it the problem that when t and u are non-zero, (these choices of) F and G must overlap in extension. The proposed rule would therefore be useless in fixing what the sum of t and u is.

This problem will be solved only by resorting instead to choices of Fx and of Gx that *are* disjoint, and for which t is the number of Fs and u is the number of Gs. In characterizing the natural numbers, even in the absence of addition and multiplication, we know that we cannot assume the existence of infinitely many concrete objects; so the numbers themselves need to be taken as objects in order to ensure the existence of indefinitely large, albeit finite, extensions of (at least, numerical) concepts. Since we have to resort to the abstract realm to ensure the existence of arbitrarily finitely many things, we may as well exploit the potential infinitude of the abstract realm in providing an account of the operations of addition and multiplication.

The current account does just that—but by exploiting the resources of orderly pairing. Regardless whether the extensions of F and of G contain any concrete objects, we can ensure respectively equinumerous *and disjoint* extensions F' and G' as follows:

$$F'(x) : \lambda(x) = 0 \wedge F(\rho(x));$$
$$G'(x) : \lambda(x) = 1 \wedge G(\rho(x)).$$

That is, the satisfiers of F' will be ordered pairs of the form $\pi(0, x)$ where x satisfies F; and the satisfiers of G' will be ordered pairs of the form $\pi(1, x)$ where x satisfies G. Since 0 is distinct from 1, F' is disjoint in extension from G', even if F and G themselves overlap. This enables one to make any individual that satisfies both F and G eligible to be reckoned 'twice over' in computing the sum of the number of Fs and the number of Gs.

In general, an arithmetical term is either of the form $\#xF(x)$, or of the form $s(t)$ for some arithmetical term t. (Note that 0 is defined to be $\#x\ \neg x = x$, so 0 qualifies as a term of the former kind.) The constructive logicist has already secured the general result that

$$\underline{n} = \#x(x < n).$$

11.2 A Second (but Still Inadequate) Stab at Rules for Addition

It may be thought that, in accounting for the conditions under which one can assert or infer statements of the form

$$t = u + v$$

where u and v are arithmetical terms in general, it should suffice to account for the conditions under which one can assert or infer statements of the form

$$t = \#xF(x) + \#xG(x).$$

Recalling that 1 is defined as $s(0)$, and that we have the theorem $\neg 0 = 1$, it might be thought that the introduction rule for addition could accordingly be framed as follows.

$$(+\text{-Intro})\ \frac{t = \#x((\lambda x = 0 \wedge F\rho x) \vee (\lambda x = 1 \wedge G\rho x))}{t = \#xF(x) + \#xG(x)}$$

The corresponding elimination rule would then simply be the converse:

$$(+\text{-Elim})\ \frac{t = \#xF(x) + \#xG(x)}{t = \#x((\lambda x = 0 \wedge F\rho x) \vee (\lambda x = 1 \wedge G\rho x))}$$

To proceed in this fashion, however, would be to lose sight of the need, on which Frege insisted so vigorously, to provide an account of the truth-conditions of an assertion of the *general* form $t = u + v$, whatever *bedeutungsvolle Eigennamen* might be put in place of u and v. In the natural-deduction setting, the Fregean logicist is obliged to provide an account of the conditions under which one may *infer to* and *from* assertions of the form $t = u + v$, whatever terms might be put in place of u and v. Among the candidate terms are the parameters that the natural-deduction theorist uses for quantificational inferences such as $\forall$-Introduction and $\exists$-Elimination within proofs. The foregoing rules do not provide such an account.

12 Rules for Addition

Regardless of the syntactic shape of the terms t, u and v—be they abstractive terms, or terms of the form $s(t)$, *or parameters*—we must specify what it is for t to be the sum of u and v. Within the confines of standard second-order logic, this can be done somewhat laboriously as follows. Note that the expression $\Phi xy[\Psi x 1\text{–}1 \Theta y]$ means that the two-place relation Φ effects a 1–1 correspondence of the Ψs with the Θs. Purely logical rules were provided in [17, pp. 276–281], for inferring to and from claims of this form.

$$\dfrac{u = \#x Fx \quad v = \#x Gx \quad \begin{array}{c} \underbrace{\overline{F'a}^{(i)},\ \overline{G'a}^{(i)}} \\ \vdots \\ \perp \end{array} \quad Rxy[Fx \text{1-1} F'y] \quad Sxy[Gx \text{1-1} G'y] \quad t = \#x(F'x \vee G'x)}{t = u + v}\ (i)$$

This rules takes into account the possibility that F and G, the concepts whose numbers u and v are to be added, might overlap in extension. When that happens, one finds *different* concepts F' and G' that are disjoint, and respectively equinumerous with F and with G, and one takes the number of the disjunctive concept $F' \vee G'$ as the required sum. This supplies the logical letter for the spirit of Frege's Section 33.

In all applications of the rule just formulated, one could always find the desired disjoint concepts F' and G' by exploiting the simple device of pairing each object in the extension of F with some object (say, 0) and pairing each object in the extension of G with some distinct object (say, 1). Thus one would have

$$F'x \ \equiv_{df} \ \lambda x = 0 \wedge F\rho x$$

$$G'x \ \equiv_{df} \ \lambda x = 1 \wedge G\rho x$$

Thus F' and G' will apply only to ordered pairs; and will be disjoint, since 0 is distinct from 1. Moreover, F' will be equinumerous with F, and G' with G. For the 1–1 mappings R and S one could take $x = \rho y$, for which one can easily prove

$$x = \rho y[Fx \text{ 1-1 } (\lambda y = 0 \wedge F\rho y)] \quad \text{and}$$

$$x = \rho y[Gx \text{ 1-1 } (\lambda y = 1 \wedge G\rho y)]$$

Thus the third, fourth and fifth subproofs for the foregoing rule would always be to hand. One might as well, therefore, cut down the postulational work and adopt the following more streamlined rule. Bear in mind that this is possible because we now have the logic of orderly pairing as part of the logical apparatus that the logicist is entitled to bring to bear on the definitional problem at hand, namely, how to set up rules that genuinely define addition of natural numbers in fully Fregean spirit. The more streamlined rule is as follows:

$$(+\text{-Intro})\ \dfrac{u = \#x Fx \quad v = \#x Gx \quad t = \#x((\lambda x = 0 \wedge F\rho x) \vee (\lambda x = 1 \wedge G\rho x))}{t = u + v}$$

Note that $F(x)$ and $G(x)$ are schematic, to be replaced, in applications of this rule, by any two formulae with exactly the variable x free. That means, in effect, that the assertability condition called for by the rule of (+-Intro) is second-order existential:

$$\frac{\exists F \exists G(u = \#x Fx \wedge v = \#x Gx \wedge t = \#x((\lambda x = 0 \wedge F\rho x) \vee (\lambda x = 1 \wedge G\rho x)))}{t = u + v}$$

Accordingly, the elimination rule takes the following form, where the predicate-parameters $\mathcal{F}$ and $\mathcal{G}$ in the subordinate proof of φ are parametric for existential elimination. That is to say, they do not occur in the premise $t = u + v$, nor in φ, nor in any assumptions, other than those indicated, on which the upper occurrence of φ depends.

$$(+\text{-Elim}) \quad \frac{t = u + v \qquad \begin{array}{c} \underbrace{\overline{u = \#x\mathcal{F}x}^{(i)}, \ \overline{v = \#x\mathcal{G}x}^{(i)}, \ \overline{t = \#x((\lambda x = 0 \wedge \mathcal{F}\rho x) \vee (\lambda x = 1 \wedge \mathcal{G}\rho x))}^{(i)}} \\ \vdots \\ \varphi \end{array}}{\varphi}\,(i)$$

The foregoing rules clearly provide a precise sense for addition on *all* cardinal numbers, both finite and infinite.

13 Rules for Multiplication

With a primitive notion of ordered pair, we can now formulate an introduction rule for multiplication as well.

$$(\times\text{-Intro}) \frac{u = \#x Fx \quad v = \#x Gx \quad t = \#x \exists y \exists z(Fy \wedge Gz \wedge x = \pi(y, z))}{t = u \times v}$$

By considerations similar to those that led us to the right elimination rule for +, we have the following elimination rule for $\times$, where the restrictions on the predicate-parameters $\mathcal{F}$ and $\mathcal{G}$ are as they were for (+-Elim):

$$(\times\text{-Elim}) \quad \frac{t = u \times v \qquad \begin{array}{c} \underbrace{\overline{u = \#x\mathcal{F}x}^{(i)}, \ \overline{v = \#x\mathcal{G}x}^{(i)}, \ \overline{t = \#x\exists y\exists z(\mathcal{F}y \wedge \mathcal{G}z \wedge x = \pi(y, z))}^{(i)}} \\ \vdots \\ \varphi \end{array}}{\varphi}\,(i)$$

The foregoing rules clearly provide a precise sense for multiplication on *all* cardinal numbers, both finite and infinite.

14 A Full Statement of the Rules of Constructive Logicism

An improved statement of the rules for #, 0 and s proposed in [17] is as follows.[6] Together with the rules stated in Section 4 for orderly pairing, and those stated in Section 12 and Section 13 for addition and multiplication respectively, they form a complete basis for a constructive logicist derivation of full Peano-Dedekind arithmetic.[7]

0-Introduction

$$\dfrac{\begin{array}{c}\underbrace{\overset{(i)}{\overline{F(a)}} \ , \ \overline{\exists ! a}^{(i)}} \\ \vdots \\ \bot \end{array}}{0 = \#x F(x)}\,(i)$$

0-Elimination

$$\dfrac{0 = \#x F(x) \quad \exists ! t \quad F(t)}{\bot}$$

#-Introduction

$$\dfrac{\#x Fx = t \quad Rxy[Fx \text{ 1–1 } Gy]}{\#x Gx = t}$$

Recall that the condition $Rxy[Fx\ 1-1\ Gy]$ is that R effects a $1-1$ correspondence of the F s with the G s.

#-Elimination

$$\dfrac{\#x Gx = t \qquad \begin{array}{c}\underbrace{\overset{(i)}{\overline{\#x Fx = t}} \ , \ \overline{Rxy[Fx \text{ 1–1 } Gy]}^{(i)}} \\ \vdots\ F,\ R \text{ parametric} \\ B \end{array}}{B}\,(i)$$

s-Introduction

$$\dfrac{\#x Fx = t \quad Rxy[Fx \text{ 1–1 } Gy, r]}{\#x Gx = st}$$

(Here, the condition $Rxy[Fx\ 1-1\ Gy, r]$ is that R effects a $1-1$ correspondence of the F s with all the G s except r. Purely logical rules were provided in [17], pp. 276–281, for inferring to and from claims of this form.)

[6] This formulation is to be found in [19], at p. 113. It is slightly more condensed than the set of rules in [17], owing to the elimination of a redundancy first remarked by Ian Rumfitt in [15].
[7] The formal derivations collectively justifying this claim are deferred to another paper.

s-Elimination (first half)

$$\dfrac{\#xGx = st \qquad \begin{matrix} \underbrace{\overline{\#xFx = t}^{(i)}\ ,\quad \overline{Rxy[Fx\ 1\text{–}1\ Gy, a]}^{(i)}} \\ \vdots\ a, F, G, R \text{ parametric} \\ B \end{matrix}}{B}\,(i)$$

s-Elimination (second half)

$$\dfrac{u = st \qquad \begin{matrix} \overline{u = \#xHx}^{(i)} \\ \vdots\ H \text{ parametric} \\ B \end{matrix}}{B}\,(i)$$

The second half of the rule of s-Elimination in effect says that terms with s dominant can only denote objects within the range of denotations of #-terms.

14.1 Commutativity and Associativity of the Two Operations

Note that the new introduction and elimination rules for $+$ and $\times$ say nothing about zero and successor. The remaining Peano–Dedekind postulates, the well-known 're-cursion equations' for $+$ and $\times$:

$$\forall x\ x + 0 = x$$
$$\forall x \forall y\ x + sy = s(x + y)$$
$$\forall x\ x \times 0 = 0$$
$$\forall x \forall y\ x \times sy = (x \times y) + x$$

will be derivable as the 'pure numerical' reflections of their respective introduction and elimination rules.

There is something intellectually satisfying in the way in which these rules for addition and for multiplication directly secure the general commutativity and asso-ciativity of the two operations.[8]

The rules of inference and the derivations of constructive logicism are purely conceptual; it would be a mistake to think that anything like Kantian intuitions, however pure, still obtrude upon the route to the basic laws of arithmetic.

[8] Once the derivations are carried out in full detail, we realize that more is required in order to establish the commutativity and associativity of addition than just the commutativity and associa-tivity of disjunction, contrary to the impression given by John Burgess in [1], at pp. 26–27.

Appendix: Proofs of (1–8)

Proof of (1).

$$\dfrac{v=\rho(t) \qquad \dfrac{u=\lambda(t) \qquad \dfrac{\dfrac{}{t=\pi(b,v)}{\scriptstyle(2)} \qquad \dfrac{\lambda\text{-I}\ \dfrac{\dfrac{}{t=\pi(u,a)}{\scriptstyle(1)}}{u=\lambda(t)} \qquad \dfrac{\dfrac{}{t=\pi(b,v)}{\scriptstyle(2)}}{b=\lambda(t)}\ \lambda\text{-I}}{u=b}\ \text{Subst.=}}{t=\pi(u,v)}\ \text{Subst.=}}{t=\pi(u,v)}\ (1)\ \lambda\text{-E}}{t=\pi(u,v)}\ (2)\ \rho\text{-E}$$

Proof of (2).

$$\lambda\text{-E}\ \dfrac{u=\lambda(t) \qquad \dfrac{\dfrac{}{t=\pi(u,a)}{\scriptstyle(1)} \qquad \dfrac{\dfrac{}{t=\pi(u,a)}{\scriptstyle(1)}}{a=\rho(t)}\ \rho\text{-I}}{t=\pi(u,\rho(t))}\ \text{Subst.=}}{t=\pi(u,\rho(t))}\ {\scriptstyle(1)}$$

The other half of (2) is proved similarly.

Proof of (3). These are special cases of the rule of denotation for functional terms. They can also, however, be derived as follows.

$$\dfrac{\exists!\pi(u,v) \qquad \dfrac{\lambda\text{-I}\ \dfrac{\dfrac{}{a=\pi(u,v)}{\scriptstyle(1)}}{u=\lambda(a)}}{\exists!u}\ \text{Rule of Denotation for Atomic Predications}}{\exists!u}\ {\scriptstyle(1)}$$

The other half of (3) can be derived similarly.

Proof of (4).

$$\dfrac{\dfrac{\dfrac{\dfrac{\pi(t,u)=\pi(v,w)}{\exists!\pi(v,w)}}{\pi(v,w)=\pi(v,w)}}{v=\lambda(\pi(v,w))} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\pi(t,u)=\pi(v,w)}{\exists!\pi(t,u)}}{\pi(t,u)=\pi(t,u)}}{t=\lambda(\pi(t,u))} \qquad \pi(t,u)=\pi(v,w)}{t=\lambda(\pi(v,w))}}{v=t}$$

The other half of (4) is proved similarly.

Proof of (5).

$$\dfrac{\dfrac{\dfrac{t=v}{\exists! t}\quad\dfrac{u=w}{\exists! u}}{\exists!\pi(t,u)}\qquad\dfrac{\dfrac{}{a=\pi(t,u)}{\scriptstyle (1)}\qquad\dfrac{\dfrac{\dfrac{}{a=\pi(t,u)}{\scriptstyle (1)}\quad t=v}{a=\pi(v,u)}\qquad u=w}{a=\pi(v,w)}}{\pi(t,u)=\pi(v,w)}}{\pi(t,u)=\pi(v,w)}{\scriptstyle (1)}$$

Our rules therefore suffice for Peano Pairing.

Proof of (6).

$$\dfrac{\exists\ \pi(x,t)=u\qquad\dfrac{\dfrac{}{\pi(a,t)=u}{\scriptstyle (1)}}{t=\rho(u)}\ \rho\text{-I}}{t=\rho(u)}{\scriptstyle (1)}$$

The other half of (6) is proved similarly.

Proof of (7).

$$\rho\text{-E}\ \dfrac{t=\rho(u)\qquad\dfrac{\dfrac{}{u=\pi(a,t)}{\scriptstyle (1)}}{\exists x\ \pi(x,t)=u}}{\exists x\ \pi(x,t)=u}{\scriptstyle (1)}$$

The other half of (7) is proved similarly.

Proof of (8).

$$\dfrac{\begin{array}{c}\exists!\lambda t\\ \text{i.e. } \exists x\ x=\lambda(t)\end{array}\qquad\dfrac{\dfrac{}{b=\lambda(t)}{\scriptstyle (2)}\qquad\dfrac{\dfrac{\dfrac{}{t=\pi(b,a)}{\scriptstyle (1)}}{a=\rho(t)}\ \rho\text{-I}}{\exists!\rho(t)}}{\exists!\rho(t)}{\scriptstyle (1)}\ \lambda\text{-E}}{\exists!\rho(t)}{\scriptstyle (2)}$$

The other half of (8) is proved similarly.

References

1. John P. Burgess. *Fixing Frege*. Princeton University Press, Princeton and Oxford, 2005.
2. Georg Cantor. *Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen*. B. G. Teubner, Leipzig, 1883.
3. Georg Cantor. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 46:481–512, 1895. English translation by Philip E. B. Jourdain in *Contributions to the Founding of the Theory of Transfinite Numbers*, 1952, Dover publications (originally Open Court, 1915), pp. 85–136.
4. Michael Dummett. *Frege: Philosophy of Mathematics*. Harvard University Press, Cambridge, Massachusetts, 1991.

5. Gottlob Frege. *Grundgesetze der Arithmetik. I. Band*. Georg Olms Verlagsbuchhandlung, Hildesheim, 1893; reprinted 1962.
6. Gottlob Frege. *Grundgesetze der Arithmetik. II. Band*. Georg Olms Verlagsbuchhandlung, Hildesheim, 1903; reprinted 1962.
7. Gottlob Frege. *The Basic Laws of Arithmetic: Exposition of the System*. Translated and edited, with an Introduction, by Montgomery Furth. University of California Press, Berkeley, Los Angeles, London, 1964.
8. Gerhard Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, I, II:176–210, 405–431, 1934, 1935. Translated as 'Investigations into Logical Deduction', in *The Collected Papers of Gerhard Gentzen*, edited by M. E. Szabo, North-Holland, Amsterdam, 1969, pp. 68–131.
9. Richard Heck, Jr. The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*. *Journal of Symbolic Logic*, 58:579–601, 1993.
10. Richard Heck, Jr. Definition by Induction in Frege's *Grundgesetze*. In William Demopoulos, editor, *Frege's Philosophy of Mathematics*, Harvard University Press, Cambridge, MA, 1995, pp. 295–333.
11. Akihiro Kanamori. The empty set, the singleton, and the ordered pair. *Bulletin of Symbolic Logic*, 9:273–288, 2003.
12. Casimir Kuratowski. Sur la notion de l'ordre dans la théorie des ensembles. *Fundamenta Mathematicae*, 2:161–171, 1921.
13. Michael Potter. *Reason's Nearest Kin: Philosophies of Arithmetic from Kant to Carnap*. Oxford University Press, Oxford, 2000.
14. Dag Prawitz. *Natural Deduction: A Proof-Theoretical Study*. Almqvist & Wiksell, Stockholm, 1965.
15. Ian Rumfitt. Frege's logicism. *Proceedings of the Aristotelian Society, Supplementary Volume*, 73:151–180, 1999.
16. Neil Tennant. *Natural Logic*. Edinburgh University Press, Edinburgh, 1978.
17. Neil Tennant. *Anti-Realism and Logic: Truth as Eternal*. Clarendon Library of Logic and Philosophy, Oxford University Press, USA, 1987.
18. Neil Tennant. *The Taming of the True*. Oxford University Press, Oxford, 1997.
19. Neil Tennant. A general theory of abstraction operators. *The Philosophical Quarterly*, 54(214):105–133, 2004.
20. Crispin Wright. *Frege's Conception of Numbers as Objects*. Aberdeen University Press, Aberdeen, 1983.

Part II
Intuitionism and Constructive Mathematics

A Constructive Version of the Lusin Separation Theorem

Peter Aczel

Abstract I state and prove a constructive version of the Lusin Separation Theorem. The classical statement of the theorem is that disjoint analytic sets are Borel separable. The definitions and results are carried out in the axiom system CZF for constructive set theory.

1 Introduction

The aim of this note is to state and prove a constructive version of the following classical result.

The Lusin Separation Theorem: *Disjoint analytic sets of Baire space are Borel separable.*

We recall the classical definitions. *Baire space* is the topological space $\mathcal{N} = {}^{\mathbb{N}}\mathbb{N}$ of infinite sequences of natural numbers, which has the product topology, $\mathbb{N}$ having the discrete topology. A subset of Baire space is *analytic* if it is a projection of a closed subset of the product space $\mathcal{N} \times \mathcal{N}$ or equivalently a continuous image of a closed subset of $\mathcal{N}$. The *Borel sets* of Baire space form the smallest σ-algebra on Baire space that includes the open sets. Here a *σ-algebra* on a set is a class of subsets of the set that contains the empty set and is closed under complements and countable unions. A pair of subsets of Baire space is *Borel separable* if there is a Borel set that includes one subset and is disjoint from the other.

The Lusin Separation Theorem has, as an immediate consequence, Suslin's fundamental classical theorem that if a subset of Baire space is both analytic and coanalytic; i.e. the complement of an analytic set, then it is Borel. In fact the converse is also classically true, so that the Borel sets are exactly the sets that are both analytic and coanalytic. According to [9] Suslin's Theorem was anounced in [10] without any proof; the first published proof being in [6], another being in [7]. It seems that

P. Aczel (✉)
Departments of Mathematics and Computer Science, Manchester University, Manchester, UK
e-mail: petera@cs.man.ac.uk

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_6, © Springer Science+Business Media B.V. 2009

the separation theorem was only established in [4]. Moschovakis, in [9], gives two proofs of the separation theorem, the first a highly non-constructive argument by contradiction, followed by a second constructive proof using Bar Induction and Bar Recursion. Such proofs were first given in [5]. The proof of the main lemma in this paper has been based on this second proof. It is interesting to note that Brouwer, in stating his bar theorem in [2] was probably inspired by the separation theorem.[1]

Our aim is to prove a version of Lusin's result in the constructive set theory **CZF**. One of the standard classical proofs, see [9], using Bar Induction and Bar Recursion would appear to be essentially constructive. But care is needed in formulating and using the assumption that the analytic subsets A_1, A_2 of Baire space, $\mathcal{N}$, are disjoint. The straightforward formulation that $A_1 \cap A_2$ is empty, may be stated as follows.

$$\neg(\exists\alpha_1 \in A_1)(\exists\alpha_2 \in A_2)[\alpha_1 = \alpha_2].$$

The equality $\alpha_1 = \alpha_2$ may be written $\neg(\exists n \in \mathbb{N})[\alpha_1 n \neq \alpha_2 n]$ so that, using constructively correct logical equivalences we may restate the disjointness of A_1, A_2 as

$$(\forall\alpha_1 \in A_1)(\forall\alpha_2 \in A_2)\ \neg\neg(\exists n \in \mathbb{N})[\alpha_1 n \neq \alpha_2 n].$$

We get a stronger notion by removing the double negation from the above. We then write that A_1, A_2 are *positively disjoint*. Wim Veldman, [12], considers that this is the natural constructive notion of disjointness. Note that the two notions become equivalent when Markov's Principle **(MP)** is assumed. This is the following principle.

Markov's Principle (MP): For each decidable subset R of $\mathbb{N}$; i.e. $(\forall n \in \mathbb{N})[(n \in R) \vee \neg(n \in R)]$,

$$\neg\neg(\exists n \in \mathbb{N})(n \in R) \ \Rightarrow\ (\exists n \in \mathbb{N})(n \in R).$$

But we do not consider **MP** to be constructively acceptable, although it is accepted by the Russian school of recursive constructive mathematics.

Even the assumption that the analytic sets are positively disjoint does not seem to be strong enough to deduce constructively, using the standard constructive forms of Bar Induction[2] and Bar Recursion, their Borel separation. Wim Veldman has shown how to overcome this problem when the analytic sets are strictly analytic, a restricted notion of analytic set introduced by Veldman [11, 12]. Here we overcome the problem in another way by strengthening the disjointness assumption even further. We will formulate a notion of *barred disjointness* for pairs of analytic sets. We will recapture a version of Veldman's result as a consequence of our main result.

---

[1] See the discussion in the introduction to the translation of [2] in [3].

[2] In the literature this standard form of Bar Induction is often refered to as BI_D.

While Bar Induction is an acceptable principle of Brouwer's Intuitionistic mathematics it is not an accepted principle of Bishop's constructive mathematics and is not provable in **CZF**. In order to avoid using Bar Induction we will strengthen the premiss of Lusin's Separation Theorem even further by using a strong point-free formulation of the disjointness property for analytic sets. This point-free notion of disjointness is equivalent to barred disjointness when Bar Induction is assumed. To compensate for the strengthening of the assumption we will also strengthen the conclusion. So we will define when analytic sets A_1, A_2 are *strongly disjoint* and when they are *strongly Borel separable* and prove the following result.

Theorem 1 (Constructive Lusin Separation Theorem) *Strongly disjoint analytic sets of Baire space are strongly Borel separable.*

In our point-free approach to the separation theorem we will use trees to represent analytic sets and Borel codes to represent Borel sets. We will define a relation $\leq$ between trees T and Borel codes b such that if $T \leq b$ then the analytic set represented by the tree T will be a subset of the Borel set represented by the code b. We will define a binary operation on trees that assigns to trees T_1, T_2 a tree $T_1 \wedge T_2$ to represent the analytic set $A_1 \cap A_2$, where A_1, A_2 are the analytic sets represented by T_1, T_2 respectively. Also we will define a unary operation on Borel codes that assigns to each Borel code b a Borel code $-b$ that represents a Borel set disjoint from the Borel set represented by b. Using these notions we can define the key concepts used in the statement of our constructive separation theorem.

Definition 2

1. Analytic sets A_1, A_2 are *strongly disjoint* if there are trees T_1, T_2, representing A_1, A_2 respectively, such that

$$T_1 \wedge T_2 \leq \Box,$$

   where $\Box$ is a Borel code for the empty Borel set.
2. Analytic sets A_1, A_2 are *strongly Borel separable* if there are trees T_1, T_2, representing A_1, A_2 respectively, such that there is a Borel code b such that

$$T_1 \leq b \text{ and } T_2 \leq -b.$$

With this definition Theorem 1 is an immediate consequence of the main lemma.

Main Lemma *If T_1, T_2 are trees such that $T_1 \wedge T_2 \leq \Box$ then there is a Borel code b such that $T_1 \leq b$ and $T_2 \leq -b$.*

Our point-free approach to Borel sets was inspired by Per Martin-Löf's constructive recursive treatment in the book [8] and the constructive approach to point-free topology developed there and in the literature on formal topology. The book formulates a point-free version of the subset relation between Borel sets on Cantor space. Here we have chosen to focus on Baire space. The book might easily have contained an extra chapter on analytic sets and the separation theorem. In fact Martin-Löf had envisioned[3] such a chapter at the time of writing his book.

A key feature of point-free topology has been the aim to prove constructive versions of classical results while avoiding any use of Bar Induction. Here we also seek to avoid Bar Induction in proving a version of a classic result of descriptive set theory.

The above theorem will be proved informally in the constructive set theory **CZF**, which we consider to be entirely compatible with Brouwer's Intuitionism. The reader is referred to [1] for an introduction to **CZF**. An important feature of our proof is the extensive use of inductive definitions of classes. One advantage of working in the set theory **CZF** is that the theory allows a flexible application of such inductive definitions.

By assuming Bar Induction we get an intuitionist separation result that assumes only that the analytic sets are barred-disjoint, a notion we introduce in Section 5. By assuming both Bar Induction and Markov's Principle we get a proof of the classical formulation of the separation theorem, and as $\mathbf{ZF} + \mathbf{DC}$ has all the theorems of $\mathbf{CZF} + \mathbf{BI} + \mathbf{MP}$ we recapture the classical result. The formal system **CZF** does not have any form of choice principle, not even countable choice, which is usually accepted in constructive mathematics. We avoid needing countable choice by working with codes of Borel sets. We prefer to avoid any form of choice whenever possible, thereby making our results more compatible with the mathematics generally true in a topos. It should be noted that when countable choice is not assumed then the standard proof that all Borel sets are analytic no longers works.

In Section 2 we present our constructive approach to the definition of the Borel sets, which uses an inductive definition of a class of codes for Borel sets, and is not essentially very different from the approaches that may be found in [8, 11, 12]. Section 3 contains a review of the **CZF** approach to inductive definitions and their application to the definition and properties of Borel sets discussed in Section 2. The notion of a tree plays an important role in the theory of analytic sets and these are discussed in Section 4. The well-founded trees are defined inductively and well-founded tree induction and well-founded tree recursion are shown to hold in **CZF**. These are variants of Bar Induction and Bar Recursion for trees that can be proved in **CZF** because the usual assumption that the tree is barred is replaced by the assumption that the tree is well-founded. The analytic sets are defined in Section 5, and four notions of disjointness for pairs of analytic sets are considered, all being equivalent if one assumes both Bar Induction and Markov's Principle. The strong inductive point-free form of Borel separation is introduced in Section 6 and the main lemma concerning trees and Borel codes is stated and proved in Section 7. Theorem 1 is an easy consequence of this main lemma. Section 8 characterises when a pair of trees represent positively disjoint analytic sets. The result is then used in Section 9 to obtain Veldman's result concerning positively disjoint strictly analytic sets as another consequence of the main lemma.

---

[3] Private Communication.

In this paper attention has been limited to the constructive analysis of Lusin's separation Theorem for Baire space only. But modern classical Descriptive Set Theory is a theory concerning Polish spaces; i.e. separable completely metrizable spaces. It is very plausible that our constructive treatment of Lusin's theorem for Baire space should carry over in a fairly routine way to Polish spaces. But we have not had time to examine this matter. Also, Lusin's Theorem is only one of a number of classical results in Descriptive Set Theory that should be amenable to a constructive treatment. In fact it should be worthwhile to develop a constructive descriptive set theory developing further the approach taken in this paper and relating it to Veldman's Intuitionistic approach. One potential application, pointed out to me by Yiannis Moschovakis is the possibility to apply results proved constructively to both the classical and recursive versions of descriptive set theory by using suitable realisability models. This remains to be studied.

I am grateful to Yiannis Moschovakis for recently drawing my attention to the Lusin Separation Theorem and its constructive character. I am grateful to Wim Veldman for drawing my attention to his work on Intuitionistic Descriptive Set Theory [11, 12], and in particular to his separation result for positively disjoint strictly analytic sets. The anonymous referee made some useful suggestions which has helped to improve the presentation of this paper.

2 Constructive Borel Sets

From now on we restrict our attention to Baire space; i.e. the space $\mathcal{N}$ of all infinite sequences of natural numbers, given the product topology where $\mathbb{N}$ is given the discrete topology. So a natural basis of clopen sets for the topology can be indexed by the set $\mathbb{N}^* = \bigcup_{n\in\mathbb{N}} \mathbb{N}^n$ of finite sequences of natural numbers. With each index $a \in \mathbb{N}^*$ of length n is associated the clopen set $G_a = \{\alpha \in \mathcal{N} \mid \overline{\alpha}n = a\}$ where if $\alpha \in \mathcal{N}$ then $\overline{\alpha}n = (\alpha 0, \ldots, \alpha(n-1))$. We may call these the *elementary clopen sets*. Note that the complement of an elementary clopen set will not be elementary. Each open set of Baire space is determined by a subset X of $\mathbb{N}^*$, and then the open set has the form

$$G_X = \{\alpha \in \mathcal{N} \mid (\exists n \in \mathbb{N})\, \overline{\alpha}n \in X\}.$$

The elementary clopen sets are not closed under taking complements. So, for our purposes a nicer basis of clopen sets uses, as indices, pairs (n, X) where $n \in \mathbb{N}$ and X is a decidable subset of $\mathbb{N}^n$. Let S be the set of such pairs. With each index $s = (n, X) \in S$ we can associate the clopen set $G_s = \{\alpha \in \mathcal{N} \mid \overline{\alpha}n \in X\}$ and its complement G_{-s}, where $-s = (n, \mathbb{N}^n - X) \in S$. We will call such sets the *simple clopen sets*. These form a base for the topology that is closed under taking complements, finite intersections and finite unions.

Classically the Borel sets can be defined to be the closure of the simple clopen sets under countable unions and intersections. Moreover every open set can be represented as a countable union of simple clopen sets and so is a Borel set. But in

CZF we cannot expect every open set to be a countable union of simple clopen sets. This is because, although the simple clopen sets form a countable set and every open set is a union of a subset of that countable set, the argument that every subset of a countable set is countable is not constructively acceptable. Let us call an open set *countably open* if it is a countable union $\bigcup_{n\in\mathbb{N}} A_n$ of an $\mathbb{N}$-indexed family $\{A_n\}_{n\in\mathbb{N}}$ of simple clopen sets A_n. Similarly we may define the *countably closed* sets to be the countable intersections $\bigcap_{n\in\mathbb{N}} A_n$ of $\mathbb{N}$-indexed families $\{A_n\}_{n\in\mathbb{N}}$ of simple clopen sets A_n.

We can get a notion of Borel set by taking the Borel sets to be obtained from the simple clopen sets by repeatedly taking countable intersections and countable unions. Because we do not want to assume countable choice we will not use that definition, but instead work with a more constructive notion by first inductively generating indices for the Borel sets.

Definition 3 The class $\mathcal{B}$ of *Borel codes* is defined to be the smallest class such that

1. $(0, s) \in \mathcal{B}$ for each index $s \in S$ for a simple clopen set,
2. If $f : \mathbb{N} \to \mathcal{B}$ then $(i, f) \in \mathcal{B}$ for $i = 1, 2$.

With each Borel code $b \in \mathcal{B}$ we associate a set $\mathbb{B}_b \subseteq \mathcal{N}$ by recursion following the inductive definition so that

1. $\mathbb{B}_b = G_s$ if $b = (0, s)$ where $s \in S$.
2. $\mathbb{B}_b = \bigcup_{n\in\mathbb{N}} \mathbb{B}_{fn}$ if $b = (1, f)$ where $f : \mathbb{N} \to \mathcal{B}$,
3. $\mathbb{B}_b = \bigcap_{n\in\mathbb{N}} \mathbb{B}_{fn}$ if $b = (2, f)$ where $f : \mathbb{N} \to \mathcal{B}$.

We define the *constructive Borel sets* to be the sets $\mathbb{B}_b$ for $b \in \mathcal{B}$.

Note that every countably closed set is Borel.

Definition 4 The duality operation $- : \mathcal{B} \to \mathcal{B}$ on Borel codes is the unique class function such that

1. $-(0, s) = (0, (n, \mathbb{N}^n - X))$ for each index $s = (n, X) \in S$.
2. $-(i, f) = (3 - i, (\lambda n \in \mathbb{N})-(fn))$ for $i = 1, 2$ and $f : \mathbb{N} \to \mathcal{B}$.

Let $\square$ be the Borel code $(0, (0, \emptyset))$. Then $\mathbb{B}_\square = \emptyset$ and $\mathbb{B}_{-\square} = \mathcal{N}$.

Proposition 5 *For all* $b \in \mathcal{B}$

1. $-b = b$*,*
2. $\mathbb{B}_b \cap \mathbb{B}_{-b} = \emptyset$*.*

Definition 6 The *complementary Borel pairs* are the pairs of Borel sets of the form

$$\mathbb{B}_b \ , \ \mathbb{B}_{-b}$$

for $b \in \mathcal{B}$.

3 Inductive Definitions in CZF

Our definition of the class $\mathcal{B}$ of Borel codes was an inductive definition. We now state the result, which can be proved in **CZF**, which justifies that inductive definition and the associated recursive definitions that assign to each Borel code b the Borel set $\mathbb{B}_b$ and the dual Borel code $-b$.

In general we take an inductive definition in **CZF** to be given as a class Φ of pairs (X, a). We call such pairs *steps*, X being the set of *premisses* of the step and a being the *conclusion* of the step. Given an inductive definition Φ we define a class Y to be Φ-*closed* if

$$X \subseteq Y \Rightarrow a \in Y$$

for every step (X, a). A proof of the following result may be found in [1].

Theorem 7 (CZF) *For each inductive definition Φ there is a (uniquely determined) smallest Φ-closed class $I(\Phi)$, the* class inductively defined by Φ. *More generally, for each class A there is a unique smallest Φ-closed class that includes the class A. We will write $I(\Phi, A)$ for this class.*

Using this result we can take the class $\mathcal{B}$ of Borel codes to be the class $I(\Phi, \{0\} \times S)$ where Φ is the inductive definition whose steps are the pairs $(\{fn \mid n \in \mathbb{N}\}, (i, f))$ for $(i, f) \in Q$ where Q is the class of pairs (i, f) where f a function with domain $\mathbb{N}$ and $i = 1, 2$. Also we can take the class function assigning the Borel set $\mathbb{B}_b$ to each Borel code $b \in \mathcal{B}$ to be the class $I(\Phi', A')$ where

$$A' = \{((0, s), G_s) \mid s \in S\},$$

and

$$\begin{aligned}\Phi' = &\{(\{(fn, gn) \mid n \in \mathbb{N}\}, ((i, f), B)) \mid \\ &(i, f) \in Q \;\;\&\;\; g : \mathbb{N} \to Pow(\mathcal{N}) \;\;\& \\ &[(i = 1 \;\&\; B = \cup_{n\in\mathbb{N}} gn) \;\vee\; (i = 2 \;\&\; B = \cap_{n\in\mathbb{N}} gn)]\}.\end{aligned}$$

Of course it is necessary to prove that $I(\Phi', A')$ is a class function $\mathcal{B} \to Pow(\mathcal{N})$ and this can be done by induction on Φ.

Next, the (graph of the) duality operator $-$ is the class $I(\Phi'', A'')$ where

$$A'' = \{((0, s), (0, -s) \mid s \in S\}$$

and

$$\Phi'' = \{(\{(fn, gn) \mid n \in \mathbb{N}\}, ((i, f), (3 - i, g))) \mid (i, f), (3 - i, g) \in Q\}.$$

Again it is necessary to prove that $I(\Phi'', A'')$ is a function $\mathcal{B} \to \mathcal{B}$ and this can be done by induction on Φ. Also Proposition 5 can be proved by induction on Φ.

4 Trees on $\mathbb{N}$

Definition 8

- If $a = (x_0, \ldots, x_{n-1}) \in \mathbb{N}^n$ and $b = (y_0, \ldots, y_{m-1}) \in \mathbb{N}^m$ then their *concatenation* is the sequence $a^\frown b = (x_0, \ldots, x_{n-1}, y_0, \ldots, y_{m-1}) \in \mathbb{N}^{n+m}$
- If $a \in \mathbb{N}^*$ then a *prefix* of a is a sequence $b \in \mathbb{N}^*$ such that $a = b^\frown c$ for some $c \in \mathbb{N}^*$.
- By a *tree* we shall always mean a decidable prefix-closed subset of $\mathbb{N}^*$; i.e. $T \subseteq \mathbb{N}^*$ is a tree if, for all $a \in \mathbb{N}^*$,
  - either $a \in T$ or $a \notin T$, and
  - if $a \in T$ then every prefix of a is also in T.
- For each tree T let

$$[T] = \{\alpha \in \mathcal{N} \mid (\forall n \in \mathbb{N})\ [\overline{\alpha}n \in T]\}.$$

Note that for each tree T the set $[T]$ is always a countably closed set.

Definition 9 A tree T is *barred* if

$$(\forall \alpha \in \mathcal{N})(\exists n \in \mathbb{N})\ [\overline{\alpha}n \notin T]$$

and is *weakly barred* if

$$(\forall \alpha \in \mathcal{N})\neg\neg(\exists n \in \mathbb{N})\ [\overline{\alpha}n \notin T].$$

Note that a tree T is weakly barred iff $[T] = \emptyset$.

We use an inductive definition to formulate a stronger point-free version of the notion of a barred tree. Let Θ be the inductive definition whose steps are the pairs $(\{a^\frown(n) \mid n \in \mathbb{N}\}, a)$ for $a \in \mathbb{N}^*$.

Definition 10 A tree T is *well-founded (wf)* if the empty sequence () is in $I(\Theta, (\mathbb{N}^* - T))$.

Proposition 11 *If T is a wf tree then* $\mathbb{N}^* \subseteq I(\Theta, (\mathbb{N}^* - T))$.

Proof Let T be a wf tree and let $I = I(\Theta, (\mathbb{N}^* - T))$. Let Y be the class $\{a \in \mathbb{N}^* \mid (\forall c \in \mathbb{N}^*)[a^\frown c \in I\}$. Then $Y \subseteq I$ and it suffices to show that $() \in Y$. In fact, as $() \in I$, it is enough to show that $I \subseteq Y$, which we do by showing that

1. $(\mathbb{N}^* - T) \subseteq Y$,
2. Y is Θ-closed.

For 1 observe that if $a \in (\mathbb{N}^* - T)$ then, as T is prefix-closed,

$$a^\frown c \in (\mathbb{N}^* - T) \subseteq I,$$

for any $c \in \mathbb{N}^*$, so that $a \in Y$.

For 2 let $(\forall n \in \mathbb{N})[a^\frown(n) \in Y]$. Then, as $Y \subseteq I$ and I is Θ-closed, $a^\frown() = a \in I$. Also, for all $n \in \mathbb{N}$ and all $c \in \mathbb{N}^*$

$$a^\frown(n)^\frown c \in I.$$

So $a^\frown e \in I$ for every $e \in \mathbb{N}^*$, as every such e is either () or else has the form $(n)^\frown c$. Thus $a \in Y$, as desired.
■

We may restate this result as the following principle.

Well-Founded Tree Induction *Let T be a wf tree. Let $Y \subseteq \mathbb{N}^*$ be a class such that*

1. $(\mathbb{N}^* - T) \subseteq Y$,
2. *Y is Θ-closed.*

Then $\mathbb{N}^* \subseteq Y$.

The following result provides an alternative characterisation of the wf trees. If T_n is a tree for each $n \in \mathbb{N}$ then let $\Sigma_{n\in\mathbb{N}} T_n$ be the tree

$$\{()\} \cup \bigcup_{n\in\mathbb{N}} \{(n)^\frown a \mid a \in T_n\}.$$

Proposition 12 *The class of wf trees is the smallest class $\mathbb{W}$ of trees such that*

1. *$\emptyset \in \mathbb{W}$,*
2. *If $T_n \in \mathbb{W}$ for $n \in \mathbb{N}$ then $\Sigma_{n\in\mathbb{N}} T_n \in \mathbb{W}$.*

Proof Let $\mathbb{W}'$ be the class of wf trees. We must show that $\mathbb{W}' = \mathbb{W}$. To show that $\mathbb{W} \subseteq \mathbb{W}'$ it suffices to observe that

1. $\emptyset$ is a wf tree,
2. If T_n is a wf tree for each $n \in \mathbb{N}$ then $\Sigma_{n\in\mathbb{N}} T_n$ is a wf tree.

For 1, observe that

$$() \in (\mathbb{N}^* - \emptyset) \subseteq I(\Theta, (\mathbb{N}^* - \emptyset)).$$

For 2 assume that, for each $n \in \mathbb{N}$, T_n is a wf tree. Let $T = \Sigma_{n\in\mathbb{N}} T_n$ and, for each $n \in \mathbb{N}$ let

$$Y_n = \{a \in \mathbb{N}^* \mid (n)^\frown a \in I(\Theta, (\mathbb{N}^* - T))\}.$$

Observe that each Y_n is Θ-closed. Also each Y_n includes $(\mathbb{N}^* - T_n)$, as

$$a \in (\mathbb{N}^* - T_n) \Rightarrow (n)^\frown a \in (\mathbb{N}^* - T) \Rightarrow (n)^\frown a \in I(\Theta, (\mathbb{N}^* - T)) \Rightarrow a \in Y_n.$$

So, for each $n \in \mathbb{N}$, by Well-Founded Tree Induction on the wf tree T_n, $\mathbb{N}^* \subseteq Y_n$. In particular, for each $n \in \mathbb{N}$, $() \in Y_n$; i.e. $(n) \in I(\Theta, (\mathbb{N}^* - T))$. So $() \in I(\Theta, (\mathbb{N}^* - T))$; i.e. the tree T is wf as desired.

To show that $\mathbb{W}' \subseteq \mathbb{W}$, let $T \in \mathbb{W}'$. So, by Proposition 11,

$$\mathbb{N}^* \subseteq I(\Theta, (\mathbb{N}^* - T)).$$

For $a \in \mathbb{N}^*$ let $T_a = \{c \in \mathbb{N}^* \mid a^\frown c \in T\}$. Clearly each T_a is a tree. As $T_{()} = T$ it suffices to show that each T_a is in $\mathbb{W}$; i.e. that $\mathbb{N}^* \subseteq Y$, where $Y = \{a \in \mathbb{N}^* \mid T_a \in \mathbb{W}\}$. Observe that (i) if $a \in (\mathbb{N}^* - T)$ then $T_a = \emptyset \in \mathbb{W}$ and (ii) if $T_{a^\frown(n)} \in \mathbb{W}$ for all $n \in \mathbb{N}$ then $T_a = \Sigma_{n \in \mathbb{N}} T_{a^\frown(n)} \in \mathbb{W}$. It follows that $\mathbb{N}^* \subseteq I(\Theta, (\mathbb{N}^* - T)) \subseteq Y$.
■

Proposition 13 *If T is a tree then*

1. *T is wf $\Rightarrow$ T is barred.*
2. *T is barred $\Rightarrow$ T is weakly barred.*

Proof For 1, by Proposition 12, it suffices to observe that $\emptyset$ is a barred tree and if T_n is a barred tree for all $n \in \mathbb{N}$ then the tree $\Sigma_{n \in \mathbb{N}} T_n$ is barred. Part 2 is trivial.
■

We are ready to formulate the principle of (decidable) Bar Induction, which is the converse of part 1 of the previous proposition.

Bar Induction *Every barred tree is wf.*

The converse to part 2 of the previous proposition seems to be less constructive. Nevertheless it is an immediate consequence of **(MP)**.

Proposition 14 (Assuming MP) *Every weakly barred tree is barred.*

In fact it is not hard to see that the statement that every weakly barred tree is barred is equivalent to (**MP**).

The following result is a variant of Bar Recursion obtained by replacing the assumption that a tree T is barred by the assumption that T is wf. Of course Bar Recursion is an immediate consequence of the theorem using Bar Induction.

Theorem 15 (Well-Founded Tree Recursion) *Let T be a wf tree. Let Y be a class. If $g : (\mathbb{N}^* - T) \to Y$ and $Q^a : {}^{\mathbb{N}}Y \to Y$ for each $a \in T$ then there is a unique $F : \mathbb{N}^* \to Y$ such that, for $a \in \mathbb{N}^*$,*

$$Fa = \begin{cases} ga & \textit{if } a \notin T \\ Q^a((\lambda n \in \mathbb{N})F(a^\frown(n))) & \textit{if } a \in T. \end{cases}$$

In fact $F = I(\Psi, g)$ where Ψ is the inductive definition with steps $(f, (a, Q^a f))$ for $a \in T$ and $f : \mathbb{N} \to Y$. It is necessary to prove that this is indeed a single valued function defined on $\mathbb{N}^*$.

5 Analytic Sets

We will use a fixed bijection $\pi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ with associated projections $\pi_1, \pi_2 : \mathbb{N} \to \mathbb{N}$ such that $\pi_i(\pi(x_1, x_2)) = x_i$ for $x_1, x_2 \in \mathbb{N}$ and $i = 1, 2$. A standard example is the bijection π with definition

$$\pi(x_1, x_2) = (x_1 + x_2)(x_1 + x_2 + 1)/2 + x_2$$

for $x_1, x_2 \in \mathbb{N}$.

We lift these functions to bijections $\mathbb{N}^n \times \mathbb{N}^n \to \mathbb{N}^n$ and associated projections as follows. If $a_1 = (x_0, \dots, x_{n-1}) \in \mathbb{N}^n$ and $a_2 = (y_0, \dots, y_{n-1}) \in \mathbb{N}^n$ then let

$$\pi(a_1, a_2) = (\pi(x_0, y_0), \dots, \pi(x_{n-1}, y_{n-1})) \in \mathbb{N}^n.$$

Also, for $i = 1, 2$ if $a = (x_0, \dots, x_{n-1}) \in \mathbb{N}^n$ then let

$$\pi_i a = (\pi_i x_0, \dots, \pi_i x_{n-1}) \in \mathbb{N}^n.$$

We also lift these functions to bijections $\mathcal{N} \times \mathcal{N} \to \mathcal{N}$ and associated projections as follows. If $\alpha_1, \alpha_2 \in \mathcal{N}$ then let $\pi(\alpha_1, \alpha_2) \in \mathcal{N}$ be given by

$$\pi(\alpha_1, \alpha_2)n = \pi(\alpha_1 n, \alpha_2 n)$$

for $n \in \mathbb{N}$. Also, for $i = 1, 2$, if $\alpha \in \mathcal{N}$ let $\pi_i\alpha \in \mathcal{N}$ be given by

$$(\pi_i\alpha)n = \pi_i(\alpha n)$$

for $n \in \mathbb{N}$.

Definition 16 A set $A \subseteq \mathcal{N}$ is defined to be *analytic* if there is a tree T such that

$$\begin{aligned} A = \pi_1[T] &= \{\pi_1\gamma \mid \gamma \in [T]\} \\ &= \{\alpha \in \mathcal{N} \mid (\exists\beta \in \mathcal{N})(\forall n \in \mathbb{N})\ \pi(\overline{\alpha}n, \overline{\beta}n) \in T\}. \end{aligned}$$

We then call T a *tree representation* of the analytic set A.

We will also need a bijection $\tau : \mathbb{N}\hat{\times}\mathbb{N} \to \mathbb{N}$ and associated projections $\tau_1, \tau_2 : \mathbb{N} \to \mathbb{N}$, where

$$\mathbb{N}\hat{\times}\mathbb{N} = \{(n_1, n_2) \in \mathbb{N} \times \mathbb{N} \mid \pi_1 n_1 = \pi_2 n_2\}.$$

We first define $\tau' : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and let τ be the restriction of τ' to the set $\mathbb{N}\hat{\times}\mathbb{N}$.

- $\tau'(n_1, n_2) = \pi(\pi_1 n_1, \pi(\pi_2 n_1, \pi_2 n_2))$, for $(n_1, n_2) \in \mathbb{N} \times \mathbb{N}$,
- $\tau_i n = \pi(\pi_1 n, \pi_i(\pi_2 n))$ for $i = 1, 2$ and $n \in \mathbb{N}$.

As with π, π_1, π_2 these functions can be lifted to each $\mathbb{N}^n$ and to $\mathcal{N}$.

Proposition 17 $\tau : \mathbb{N}\hat{\times}\mathbb{N} \to \mathbb{N}$ *is a bijection with projections* $\tau_1, \tau_2 : \mathbb{N} \to \mathbb{N}$.

Proof Routine computations using the definitions show that, for $n, n_1, n_2 \in \mathbb{N}$,

$$[(n_1, n_2) \in \mathbb{N}\hat{\times}\mathbb{N} \;\;\&\;\; n = \tau(n_1, n_2)] \quad \Leftrightarrow \quad [n_1 = \tau_1 n \;\;\&\;\; n_2 = \tau_2 n].$$

■

Proposition 18 *Let* T_1, T_2 *be trees representing the analytic sets* A_1, A_2. *Then*

$$T_1 \wedge T_2 = \{a \in \mathbb{N}^* \mid \tau_1 a \in T_1 \;\&\; \tau_2 a \in T_2\}$$

is a tree that represents the analytic set $A_1 \cap A_2 = \pi_1[T_1 \wedge T_2]$. *Moreover* A_1, A_2 *are disjoint iff* $T_1 \wedge T_2$ *is weakly barred.*

Proof That $T_1 \wedge T_2$ is decidable and prefix-closed follows easily from the corresponding properties of T_1 and T_2.

Next we must show that

$$\gamma \in \pi_1[T_1 \wedge T_2] \quad \Leftrightarrow \quad \gamma \in \pi_1[T_1] \cap \pi_1[T_2].$$

Assuming the left hand side, $\gamma = \pi_1\alpha$ for some $\alpha \in [T_1 \wedge T_2]$ so that, if $\alpha_i = \tau_i\alpha$ then $\alpha_i \in [T_i]$ and $\gamma = \pi_1\alpha = \pi_1\alpha_i \in \pi_1[T_i]$ for $i = 1, 2$, giving us the right hand side.

Assuming the right hand side, there are $\alpha_1, \alpha_2 \in \mathcal{N}$ such that $\gamma = \pi_1\alpha_1 = \pi_1\alpha_2$ and for all $n \in \mathbb{N}$,

$$\overline{\alpha}_1 n \in T_1 \text{ and } \overline{\alpha}_2 n \in T_2.$$

We may define $\alpha = \tau(\alpha_1, \alpha_2)$ and observe that $\gamma = \pi_1\alpha$ and $\tau_i\alpha = \alpha_i \in [T_i]$ for $i = 1, 2$; i.e. the left hand side.

Finally,

$$\begin{aligned} A_1, A_2 \text{ are disjoint} &\Leftrightarrow \pi_1[T_1 \wedge T_2] = \emptyset \\ &\Leftrightarrow [T_1 \wedge T_2] = \emptyset \\ &\Leftrightarrow (\forall \alpha \in \mathcal{N})\neg\neg(\exists n \in \mathbb{N})[\overline{\alpha} n \notin T_1 \wedge T_2] \\ &\Leftrightarrow T_1 \wedge T_2 \text{ is weakly barred.} \end{aligned}$$

■

Definition 19 Analytic sets A_1, A_2 are defined to be *strongly disjoint (barred-disjoint)* if trees T_1, T_2 representing them can be chosen such that $T_1 \wedge T_2$ is wf (barred).

Proposition 20 *For analytic sets A_1, A_2 we have the implications*

$$D1 \Rightarrow D2 \Rightarrow D3 \Rightarrow D4$$

where

D1: A_1, A_2 *are strongly disjoint,*
D2: A_1, A_2 *are barred disjoint,*
D3: A_1, A_2 *are positively disjoint,*
D4: A_1, A_2 *are disjoint,*

Moreover, assuming (**BI**)*,* (**D1**) *and* (**D2**) *are equivalent and, assuming* (**MP**)*,* (**D2**), (**D3**) *and* (**D4**) *are equivalent.*

Proof

D1$\Rightarrow$ **D2** By part 1 of Proposition 13.
D2$\Rightarrow$ **D3** Let A_1, A_2 be barred-disjoint analytic sets. So there are trees T_1, T_2 such that $A_i = \pi_1[T_i]$ for $i = 1, 2$ such that $T_1 \wedge T_2$ is barred; i.e. for all $\alpha \in \mathcal{N}$ there is $n \in \mathbb{N}$ such that $\overline{\alpha}n \notin T_1 \wedge T_2$. Now let $\alpha_i \in [T_i]$ for $i = 1, 2$ and let $\alpha = \tau'(\alpha_1, \alpha_2)$. So there is $n \in \mathbb{N}$ such that $\overline{\alpha}n \notin T_1 \wedge T_2$; i.e.

$$[\tau_1(\overline{\alpha}n) \notin T_1] \text{ or } [\tau_2(\overline{\alpha}n) \notin T_2].$$

Note that, if $\pi_1(\overline{\alpha}_1 n) = \pi_1(\overline{\alpha}_2 n)$ then

$$\overline{\alpha}_i n = \tau_i(\overline{\alpha}n) \notin T_i$$

for $i = 1, 2$. So

$$[\pi_1(\overline{\alpha}_1 n) = \pi_1(\overline{\alpha}_2 n)] \;\Rightarrow\; [\overline{\alpha}_1 n \notin T_1 \text{ or } \overline{\alpha}_2 n \notin T_2]$$

and hence

$$[\overline{\alpha}_1 n \in T_1 \text{ and } \overline{\alpha}_2 n \in T_2] \;\Rightarrow\; [\pi_1(\overline{\alpha}_1 n) \neq \pi_1(\overline{\alpha}_2 n)].$$

As $\alpha_i \in [T_i]$ for $i = 1, 2$,

$$[\overline{\alpha}_1 n \in T_1 \text{ and } \overline{\alpha}_2 n \in T_2]$$

so that $[\pi_1(\overline{\alpha}_1 n) \neq \pi_1(\overline{\alpha}_2 n)]$.
We have shown that

$$(\forall \alpha_1 \in [T_1])(\forall \alpha_2 \in [T_2])(\exists n \in \mathbb{N})[(\pi_1(\overline{\alpha}_1 n) \neq \pi_1(\overline{\alpha}_2 n)].$$

It follows that

$$(\forall \beta_1 \in A_1)(\forall \beta_2 \in A_2)(\exists n \in \mathbb{N})[\overline{\beta}_1 n \neq \overline{\beta}_2 n];$$

i.e. A_1, A_2 are positively disjoint.

D3⇒ D4 Trivial.

The final assertion about the equivalences first assuming (**BI**) and then assuming (**MP**) follows because (**BI**) expresses that every barred tree is wf and (**MP**) implies that every weakly barred tree is barred.

■

6 Strong Borel Separation

Recall that, classically, analytic sets A_1, A_2 are Borel separable if there is a Borel set B such that $A_1 \subseteq B$ and $A_2 \subseteq \mathcal{N} - B$, where the complement of B, $\mathcal{N} - B$, is a Borel set. In our constructive context we do not know that $\mathcal{N} - B$ is itself a Borel set. Instead we will use complementary pairs, $\mathbb{B}_b, \mathbb{B}_{-b}$ of Borel sets determined by a Borel code $b \in \mathcal{B}$. In fact we will work with a strengthened notion of Borel Separation obtained by using an inductively defined point-free relation, '$T \leq b$', between tree representations T of analytic sets and Borel codes $b \in \mathcal{B}$, instead of the relation '$A \subseteq B$' between analytic sets A and Borel sets B.

Given a tree T let

$$\mathcal{G}_T = \{(a, b) \in \mathbb{N}^* \times \mathcal{B} \mid A_a \subseteq \mathbb{B}_b\}$$

where $A_a = \pi_1([T] \cap G_a)$. Note that $A_{()} = \pi_1[T]$ is the analytic set with tree representation T and, in general, for $a \in \mathbb{N}^*$, $A_a = \pi_1[T_a]$ is the analytic set with tree representation

$$T_a = \{a' \in T \mid a' \leq a \ \vee \ a \leq a'\}.$$

So $(a, b) \in \mathcal{G}_T$ iff the analytic set A_a is a subset of the Borel set $\mathbb{B}_b$. Our aim now is to give an inductive definition

$$\mathcal{F}_T = I(\Phi, Y_T),$$

of a subclass $\mathcal{F}_T$ of the set $\mathcal{G}_T$ so that we can then define the desired relation between tree-codes T and Borel codes b as follows.

$$T \leq b \ \Leftrightarrow \ ((), b) \in \mathcal{F}_T.$$

We first define the base set Y_T of the inductive definition and show that $Y_T \subseteq \mathcal{G}_T$.

Definition 21 For each tree T let

$$Y_T = \{(a, (0, (n, X))) \mid a \in \mathbb{N}^*, (n, X) \in S \text{ and } [a \notin T \ \vee \ a \in \hat{X}]\},$$

where, for each $(n, X) \in S$,

$$\hat{X} = \{a \in \mathbb{N}^* \mid \exists a' \leq a \ \pi_1 a' \in X\}.$$

Proposition 22 $Y_T \subseteq \mathcal{G}_T$.

Proof Let $(a, (0, (n, X))) \in Y_T$. So $a \in \mathbb{N}^*$, $(n, X) \in S$ and either (i) $a \notin T$ or (ii) $a \in \hat{X}$.

If (i) then $[T] \cap G_a = \emptyset$ so that $A_a = \pi_1 \emptyset = \emptyset \subseteq \mathbb{B}_b$.

If (ii) then $a \geq a'$ for some $a' \in \mathbb{N}^*$ such that $\pi_1 a' \in X$. So, as $\mathbb{B}_b = \bigcup_{a \in X} G_a$,

$$\begin{aligned} \beta \in [T] \cap G_a &\Rightarrow \beta \in G_{a'} \\ &\Rightarrow \pi_1 \beta \in G_{\pi_1 a'} \subseteq \mathbb{B}_b. \end{aligned}$$

It follows that

$$\begin{aligned} \alpha \in A_a &\Rightarrow \alpha = \pi_1 \beta \text{ for some } \beta \in [T] \cap G_a \\ &\Rightarrow \alpha \in \mathbb{B}_b. \end{aligned}$$

In either case $(a, b) \in \mathcal{G}_T$.

■

We now turn to the definition of the class Φ of steps of the inductive definition of $\mathcal{F}_T$ and the proof that $\mathcal{G}_T$ is Φ-closed. Given $a \in \mathbb{N}^*$ and $b \in \mathcal{B}$ let

$$X_b^0(a) = \{(a^\frown(m), b) \mid m \in \mathbb{N}\}.$$

Also, if $f : \mathbb{N} \to \mathcal{B}$ let

$$X_{f,n}^1(a) = \{(a, f(n))\} \text{ for each } n \in \mathbb{N}$$

and

$$X_f^2(a) = \{(a, f(m)) \mid m \in \mathbb{N}\}.$$

Definition 23 Let Φ be the class of all pairs $(X, (a, b))$ such that $(a, b) \in \mathbb{N}^* \times \mathcal{B}$ and either $X = X_b^0(a)$ or $b = (i, f)$, with $i \in \{1, 2\}$ and $f : \mathbb{N} \to \mathcal{B}$, and if $i = 1$ then $X = X_{f,n}^1(a)$ for some $n \in \mathbb{N}$ and if $i = 2$ then $X = X_f^2(a)$.

Proposition 24 *$\mathcal{G}_T$ is Φ-closed.*

Proof Let $(X, (a, b)) \in \Phi$ such that $X \subseteq \mathcal{G}_T$. Then $(a, b) \in \mathbb{N}^* \times \mathcal{B}$ and we must show that $(a, b) \in \mathcal{G}_T$. There are three cases.

$X = X^0_b(a)$: By assumption, $(a^\frown(m), b) \in \mathcal{G}_T$ for all $m \in \mathbb{N}$; i.e.

$$A_{a^\frown(m)} \subseteq \mathbb{B}_b \text{ for all } m \in \mathbb{N}.$$

Observe that

$$A_a = \bigcup_{m \in \mathbb{N}} A_{a^\frown(m)}.$$

It follows that $A_a \subseteq \mathbb{B}_b$.

$X = X^1_{f,n}(a)$, with $n \in \mathbb{N}$, $b = (1, f)$ and $f : \mathbb{N} \to \mathcal{B}$:
In this case, as $(a, f(n)) \in \mathcal{G}_T$,

$$A_a \subseteq \mathbb{B}_{f(n)} \subseteq \bigcup_{m \in \mathbb{N}} \mathbb{B}_{f(m)} = \mathbb{B}_b.$$

$X = X^2_f(a)$, with $b = (2, f)$ and $f : \mathbb{N} \to \mathcal{B}$:
As $(a, f(m)) \in \mathcal{G}_T$ for all $m \in \mathbb{N}$,

$$A_a \subseteq \bigcap_{m \in \mathbb{N}} \mathbb{B}_{f(m)} = \mathbb{B}_b.$$

In all three cases we have shown that $(a, b) \in \mathcal{G}_T$.
■

By Propositions 22 and 24 we get the following result.

Proposition 25 $\mathcal{F}_T \subseteq \mathcal{G}_T$.

Recall the definition $T \leq b \;\Leftrightarrow\; ((), b) \in \mathcal{F}_T$. We get the following corollary, using Definition 10 for part 2.

Corollary 26 *Let T be a tree.*

1. *If $b \in \mathcal{B}$ such that $T \leq b$ then $\pi_1[T] \subseteq \mathbb{B}_b$.*
2. *The tree T is wf iff $T \leq \Box$.*

Definition 27 Analytic sets A_1, A_2 are *strongly Borel separable* if trees T_1, T_2 representing them and $b \in \mathcal{B}$ can be chosen such that $T_1 \leq b$ and $T_2 \leq -b$.

For each $b \in \mathcal{B}$ let

$$\mathcal{F}_T(b) = \{a \in \mathbb{N}^* \mid (a, b) \in \mathcal{F}_T\}.$$

The following proposition is just a reformulation of the fact that, for any tree T, the class $\mathcal{F}_T$ is Φ-closed.

Proposition 28 *For all $a \in \mathbb{N}^*$, $b \in \mathcal{B}$ and $f : \mathbb{N} \to \mathcal{B}$,*

F0 $(\forall m \in \mathbb{N})[a^\frown(m) \in \mathcal{F}_T(b)] \;\Rightarrow\; a \in \mathcal{F}_T(b)$,

F1 $(\exists n \in \mathbb{N})[a \in \mathcal{F}_T(f(n))] \;\Rightarrow\; a \in \mathcal{F}_T((1, f))$,

F2 $(\forall n \in \mathbb{N})[a \in \mathcal{F}_T(f(n))] \;\Rightarrow\; a \in \mathcal{F}_T((2, f))$.

The next result expresses the crucial idea behind the constructive proof of the Lusin theorem.

Proposition 29 *Given trees T_1, T_2, if $a_1, a_2 \in \mathbb{N}^*$ and $h : \mathbb{N} \times \mathbb{N} \to \mathcal{B}$ such that*

$$(*) \quad a_1{}^\frown(n_1) \in \mathcal{F}_{T_1}(h(n_1, n_2)) \;\;\&\;\; a_2{}^\frown(n_2) \in \mathcal{F}_{T_2}(-h(n_1, n_2))$$

for all $n_1, n_2 \in \mathbb{N}$ then

$$a_1 \in \mathcal{F}_{T_1}(b) \;\;\&\;\; a_2 \in \mathcal{F}_{T_2}(-b)$$

where

$$b = (1, (\lambda n_1 \in \mathbb{N})(2, (\lambda n_2 \in \mathbb{N})h(n_1, n_2))).$$

Proof Let $f = (\lambda k \in \mathbb{N})(2, h_k)$ where $h_k = (\lambda m \in \mathbb{N})h(k, m)$. Then $b = (1, f)$.
By (*), F2) and F1), for all $n_1 \in \mathbb{N}$,

$$a_1{}^\frown(n_1) \in \bigcap_{m \in \mathbb{N}} \mathcal{F}_{T_1}(h_{n_1}(m)) \subseteq \mathcal{F}_{T_1}((2, h_{n_1})) = \mathcal{F}_{T_1}(f(n_1)) \subseteq \mathcal{F}_{T_1}(b),$$

So by F0), $a_1 \in \mathcal{F}_{T_1}(b)$.
Let $f^- = (\lambda k \in \mathbb{N})(1, h_k^-)$ where $h_k^- = (\lambda m \in \mathbb{N}) - h(k, m)$. Then $b = (2, f^-)$.
By (*) and F1), for all $n_1, n_2 \in \mathbb{N}$,

$$a_2{}^\frown(n_2) \in \mathcal{F}_{T_2}(h_{n_1}^-(n_2)) \subseteq \mathcal{F}_{T_2}((1, h_{n_1}^-)) = \mathcal{F}_{T_2}(f^-(n_1)).$$

So, by F0),

$$a_2 \in \mathcal{F}_{T_2}(f^-(n_1)) \text{ for all } n_1 \in \mathbb{N},$$

so that, by F2),

$$a_2 \in \mathcal{F}_{T_2}((2, f^-)) = \mathcal{F}_{T_2}(-b).$$

7 The Main Lemma

Theorem 1 is an easy consequence of the following point-free result.

Lemma 30 (Main Lemma) *If T_1, T_2 are trees such that $T_1 \wedge T_2 \leq \Box$ then $T_1 \leq b$ and $T_2 \leq -b$ for some $b \in \mathcal{B}$.*

Note: The converse result, that if $T_1 \leq b$ and $T_2 \leq -b$ then $T_1 \wedge T_2 \leq \Box$, is plausible, but has not been proved yet.

We now start the proof of the Main Lemma. We will need functions $h^a : \mathbb{N} \times \mathbb{N} \to \mathcal{B}$ for $a \in \mathbb{N}^*$, given by

$$h^a(n_1, n_2) = \begin{cases} q^a n_1 & \text{if } \pi_1 n_1 \neq \pi_1 n_2 \\ f(\tau(n_1, n_2)) & \text{if } \pi_1 n_1 = \pi_1 n_2 \end{cases}$$

for $n_1, n_2 \in \mathbb{N}$ where, if $a \in \mathbb{N}^m$ then $q^a : \mathbb{N} \to \mathcal{B}$ is given by

$$q^a n = (0, (m+1, X_n))$$

for $n \in \mathbb{N}$, where $X_n = \{c^\frown(\pi_1 n) \mid c \in \mathbb{N}^m\}$.

Lemma 31 *For each tree T, if $a \in \mathbb{N}^*$ and $n, n' \in \mathbb{N}$ such that $\pi_1 n \neq \pi_1 n'$ then*

$$a^\frown(n) \;\in\; \mathcal{F}_T(q^a n) \;\subseteq\; \mathcal{F}_T(-q^a n').$$

Proof Let $a \in \mathbb{N}^m$. As $\pi_1 a \in \mathbb{N}^m$,

$$\pi_1(a^\frown(n)) = \pi_1 a\ ^\frown(\pi_1 n) \in X_n$$

so that $a^\frown(n) \in \hat{X}_n \subseteq \mathcal{F}_T(q^a n)$.

Now let $n, n' \in \mathbb{N}$ such that $\pi_1 n \neq \pi_1 n'$. It remains to show that $\mathcal{F}_T(q^a n) \subseteq \mathcal{F}_T(-q^a n')$. Observe that

$$\mathcal{F}_T(q^a n) = (\mathbb{N}^* - T) \cup \hat{X}_n,$$
$$\mathcal{F}_T(-q^a n') = (\mathbb{N}^* - T) \cup \hat{Y}_{n'},$$

where $Y_{n'} = (\mathbb{N}^{m+1} - X_{n'})$. So it suffices to show that

$$(*) \qquad \hat{X}_n \subseteq \hat{Y}_{n'}.$$

Observe that

$$a \in \hat{X}_n \;\Leftrightarrow\; (\exists c \in \mathbb{N}^m)(\exists n_0 \in \mathbb{N})(\exists d \in \mathbb{N}^*)[a = c^\frown(n_0)^\frown d \text{ and } \pi_1 n_0 = \pi_1 n],$$
$$a \in \hat{Y}_{n'} \;\Leftrightarrow\; (\exists c \in \mathbb{N}^m)(\exists n_0' \in \mathbb{N})(\exists d \in \mathbb{N}^*)[a = c^\frown(n_0')^\frown d \text{ and } \pi_1 n_0' \neq \pi_1 n'].$$

As $\pi_1 n \neq \pi_1 n'$ we get $(*)$ using $n_0' = n_0$.

■

Let T_1, T_2 be trees such that $T = T_1 \wedge T_2 \leq$ so that T is wf. We define $F : \mathbb{N}^* \to \mathcal{B}$ by Well-Founded Tree Recursion, Theorem 15, on T so that for $a \in \mathbb{N}^*$

$$Fa = \begin{cases} ga & \text{if } a \notin T \\ Q^a((\lambda n \in \mathbb{N})F(a^\frown(n))) & \text{if } a \in T \end{cases}$$

where, if $a \notin T$ then

$$ga = \begin{cases} \Box & \text{if } \tau_1 a \notin T_1 \\ -\Box & \text{if } \tau_1 a \in T_1 \end{cases}$$

and, if $a \in T$ and $f : \mathbb{N} \to \mathcal{B}$ then

$$Q^a f = (1, (\lambda n \in \mathbb{N})(2, (\lambda m \in \mathbb{N})h^a(n, m)))$$

To complete the proof of the main lemma it is enough to apply the following lemma with $a = ()$ and put $b = F()$.

Lemma 32 *For all* $a \in \mathbb{N}^*$

$$(*) \quad \tau_1 a \in \mathcal{F}_{T_1}(Fa) \quad \textit{and} \quad \tau_2 a \in \mathcal{F}_{T_2}(-Fa).$$

Proof Let Y be the class of $a \in \mathbb{N}^*$ such that $(*)$. By Well-Founded Tree Induction on the wf tree T it suffices to show that

1. $(\mathbb{N}^* - T) \subseteq Y$,
2. Y is Θ-closed

For 1: Let $a \in (\mathbb{N}^* - T)$ so that $a \notin T$ and hence $Fa = ga$ and either $\tau_1 a \notin T_1$ or $\tau_2 a \notin T_2$.

Case 1 $\tau_1 a \notin T_1$: As $\tau_1 a \in \mathbb{N}^* - T_1$ we have $Fa = ga = \Box$ so that $\tau_1 a \in (\mathbb{N}^* - T_1) \subseteq \mathcal{F}_{T_1}(Fa)$. Observe that, as $\{\hat{()}\} = \mathbb{N}^*$,

$$\mathcal{F}_{T_2}(-Fa) = \mathcal{F}_{T_2}(-\Box) = (\mathbb{N}^* - T_2) \cup \{\hat{()}\} = \mathbb{N}^*.$$

It follows that $\tau_2 a \in \mathcal{F}_{T_2}(-Fa)$.

Case 2 $\tau_1 a \in T_1$ & $\tau_2 a \notin T_2$: In this case $-ga = \Box$ and we can argue as in case 1 interchanging the roles of the subscripts 1,2.

For 2: Let $a \in \mathbb{N}^*$ such that $a^\frown(n) \in Y$ for all $n \in \mathbb{N}$. We want to show that $a \in Y$. By 1 we may assume that $a \in T$ so that $Fa = Q^a f$ where $f = (\lambda n \in \mathbb{N})\ F(a^\frown(n))$. Let $a_1 = \tau_1 a$, $a_2 = \tau_2 a$. By our initial assumption that $a^\frown(n) \in Y$ for all $n \in \mathbb{N}$ we get that if $n_1, n_2 \in \mathbb{N}$ such that $\pi_1 n_1 = \pi_1 n_2$ then

$$a_1{}^\frown(n_1) \in \mathcal{F}_{T_1}(f(\tau(n_1, n_2))) \quad \& \quad a_2{}^\frown(n_2) \in \mathcal{F}_{T_2}(-f(\tau(n_1, n_2))).$$

Also observe that, if $n_1, n_2 \in \mathbb{N}$ such that $\pi_1 n_1 \neq \pi_1 n_2$ then, by Lemma 31 below,

$$a_1 {}^\frown(n_1) \in \mathcal{F}_{T_1}(q^a n_1) \quad \& \quad a_2 {}^\frown(n_2) \in \mathcal{F}_{T_2}(-q^a n_1).$$

It follows that for all $n_1, n_2 \in \mathbb{N}$

$$a_1 {}^\frown(n_1) \in \mathcal{F}_{T_1}(h^a(n_1, n_2)) \quad \& \quad a_2 {}^\frown(n_2) \in \mathcal{F}_{T_2}(-h^a(n_1, n_2))$$

so that, by Proposition 29, $a \in Y$.
■

8 Positive Disjointness

If T_1, T_2 are trees representing the analytic sets A_1, A_2 then we may characterise that A_1, A_2 are positively disjoint, as defined in Section 1, in terms of a relative notion of barred tree as follows. When **MP** is assumed then the relative notion of barred subtree is equivalent to the unrelativised notion of barred tree.

Definition 33 If T is a tree then a subtree T' is a *barred subtree of* T if

$$(\forall \alpha \in [T])(\exists n \in \mathbb{N})\ \overline{\alpha} n \notin T'.$$

If T_1, T_2 are trees then let $T_1 \times T_2$ be the tree

$$\{a \in \mathbb{N}^* \mid \pi_1 a \in T_1 \ \& \ \pi_2 a \in T_2\}$$

and let $T_1 \hat{\times} T_2$ be the subtree

$$\{a \in T_1 \times T_2 \mid \pi_1(\pi_1 a) = \pi_1(\pi_2 a)\}.$$

Proposition 34 *If* T_1, T_2 *are trees representing the analytic sets* A_1, A_2 *then* A_1, A_2 *are positively disjoint analytic sets iff* $T_1 \hat{\times} T_2$ *is a barred subtree of the tree* $T_1 \times T_2$.

Proof Let $A_i = \pi_1[T_i]$ where T_i is a tree, for $i = 1, 2$. Note that

$$\begin{aligned} \gamma \in [T_1 \times T_2] \ &\Leftrightarrow \ (\forall n \in \mathbb{N})[\pi_1(\overline{\gamma} n) \in T_1 \ \& \ \pi_2(\overline{\gamma} n) \in T_2] \\ &\Leftrightarrow \ \pi_1 \gamma \in [T_1] \ \& \ \pi_2 \gamma \in [T_2]. \end{aligned}$$

So, A_1, A_2 are positively disjoint

$$\begin{aligned}
&\Leftrightarrow\ (\forall\alpha_1 \in A_1)(\forall\alpha_2 \in A_2)(\exists n \in \mathbb{N})[\alpha_1 n \neq \alpha_2 n] \\
&\Leftrightarrow\ (\forall\gamma_1 \in [T_1])(\forall\gamma_2 \in [T_2])(\exists n \in \mathbb{N})[\pi_1(\gamma_1 n) \neq \pi_1(\gamma_2 n)] \\
&\Leftrightarrow\ (\forall\gamma \in [T_1 \times T_2])(\exists n \in \mathbb{N})[\pi_1(\pi_1(\gamma n)) \neq \pi_1(\pi_2(\gamma n))] \\
&\Leftrightarrow\ (\forall\gamma \in [T_1 \times T_2])(\exists n \in \mathbb{N})[\pi_1(\pi_1(\overline{\gamma} n)) \neq \pi_1(\pi_2(\overline{\gamma} n))] \\
&\Leftrightarrow\ (\forall\gamma \in [T_1 \times T_2])(\exists n \in \mathbb{N})[\overline{\gamma} n \notin T_1 \hat{\times} T_2] \\
&\Leftrightarrow\ T_1 \hat{\times} T_2 \text{ is a barred subtree of } T_1 \times T_2.
\end{aligned}$$

■

Proposition 35 (Assuming MP) *If T' is a subtree of a tree T then T' is a barred subtree of T iff T' is a barred tree.*

Proof The implication from right to left is trivial. For the other direction let T' be a barred subtree of T; i.e. $T' \subseteq T$ such that, for all $\alpha \in \mathcal{N}$,

$$(*)\quad (\forall n \in \mathbb{N})[\overline{\alpha} n \in T] \Rightarrow (\exists n \in \mathbb{N})[\overline{\alpha} n \notin T'].$$

Given $\alpha \in \mathcal{N}$ we must show that $(\exists n \in \mathbb{N})[\overline{\alpha} \notin T']$. We have

$$\begin{aligned}
\neg(\exists n \in \mathbb{N})[\overline{\alpha} n \notin T'] &\Rightarrow (\forall n \in \mathbb{N})[\overline{\alpha} n \in T'], && \text{as } T' \text{ is a decidable subset of } \mathbb{N}^*, \\
&\Rightarrow (\forall n \in \mathbb{N})[\overline{\alpha} n \in T], && \text{as } T' \subseteq T, \\
&\Rightarrow (\exists n \in \mathbb{N})[\overline{\alpha} n \notin T'], && \text{by } (*).
\end{aligned}$$

Hence $\neg\neg(\exists n \in \mathbb{N})[\overline{\alpha} n \notin T']$. So, by **MP**, $(\exists n \in \mathbb{N})[\overline{\alpha} n \notin T']$.
■

Corollary 36 (Assuming MP) *If T_1, T_2 are tree representations of the analytic sets A_1, A_2 respectively then A_1, A_2 are positively disjoint iff the tree $T_1 \hat{\times} T_2$ is (weakly) barred.*

9 Strictly Analytic Sets

Using Bar Induction we get the following point-set Corollary of Theorem 1.

Theorem 37 (Assuming BI) *Barred-disjoint analytic sets are strongly Borel separable and hence Borel separable.*

If we also assume Markov's Principle then we regain (a slight strengthening of) the classical result.

Theorem 38 (Assuming both BI and MP) *Disjoint analytic sets are strongly Borel separable.*

Note that every theorem of **CZF** + **BI** + **MP** is a theorem of **ZF** + **DC**.

We apply Theorem 37 to get a version, in our setting, of a result of Wim Veldman, see Theorem 9.2 of [12].

Definition 39 A tree T is a *spread tree* if $() \in T$ and

$$(\forall a \in T)(\exists n \in \mathbb{N})\ a^\frown(n) \in T.$$

An analytic set is *strictly analytic* if it can be represented by a spread tree.

Note that an analytic set need not be strictly analytic, as the empty set is analytic but any strictly analytic set is a continuous image of the whole of Baire space and so is always an inhabited set. A construction for the continuous function is given in the proof of part (1) of Lemma 42 below.

Theorem 40 (Assuming BI) *Positively disjoint strictly analytic sets are strongly Borel separable.*

It is a consequence of Theorem 37 and the following result.

Theorem 41 *Positively disjoint strictly analytic sets are barred disjoint.*

Proof By Proposition 34 this is a consequence of the following lemma.

Lemma 42

1. *If T is a spread tree then*

$$(\forall \alpha \in \mathcal{N})(\exists \beta \in \mathcal{N})(\forall n \in \mathbb{N})\ [\overline{\beta}n \in T \ \&\ (\overline{\alpha}n \in T \Rightarrow \overline{\alpha}n = \overline{\beta}n)].$$

2. *If T_1, T_2 are spread trees such that $T_1 \hat{\times} T_2$ is a barred subtree of $T_1 \times T_2$ then $T_1 \hat{\times} T_2$ is a barred tree.*
3. *For all trees T_1, T_2, the tree $T_1 \wedge T_2$ is barred iff $T_1 \hat{\times} T_2$ is barred.*

Proof

1. Given $\alpha \in \mathcal{N}$, define $\beta \in [T]$ by primitive recursion as follows. For each $n \in \mathbb{N}$ let $\beta n = \alpha n$ if $\overline{\alpha}(n+1) \in T$. If $\overline{\alpha}(n+1) \notin T$ then let βn be the least $j \in \mathbb{N}$ such that $\overline{\beta}n^\frown(j) \in T$. Such a j will always exist as $\overline{\beta}n \in T$ and T is a spread tree.
2. Let T_1, T_2 be spread trees such that $T_1 \hat{\times} T_2$ is a barred subtree of $T_1 \times T_2$. Given $\alpha_1, \alpha_2 \in \mathcal{N}$ choose $\beta_1, \beta_2 \in \mathcal{N}$ by part (1), such that for all $n \in \mathbb{N}$, $\overline{\beta_1}n \in T_1$, $\overline{\beta_2}n \in T_2$, and

$$(\overline{\alpha_1}n \in T_1 \Rightarrow \overline{\alpha_1}n = \overline{\beta_1}n) \text{ and } (\overline{\alpha_2}n \in T_1 \Rightarrow \overline{\alpha_2}n = \overline{\beta_2}n)$$

As $\beta_1 \in [T_1]$ and $\beta_2 \in T_2$, $\pi(\beta_1, \beta_2) \in [T_1 \times T_2]$ so that there is $n \in \mathbb{N}$ such that $\pi(\overline{\beta_1}n, \overline{\beta_2}n) \notin T_1 \hat{\times} T_2$ and hence $\pi_1(\overline{\beta_1}n) \neq \pi_1(\overline{\beta_2}n)$. It follows that if $\overline{\alpha_1}n \in T_1$ and $\overline{\alpha_2}n \in T_2$ then $\overline{\alpha_1}n = \overline{\beta_1}n$ and $\overline{\alpha_2}n = \overline{\beta}n$ so that $\pi_1(\overline{\alpha_1}n) \neq \pi_1(\overline{\alpha_2}n)$. Thus $\overline{\pi(\alpha_1, \alpha_2)}n \notin T_1 \hat{\times} T_2$.
We have shown that $T_1 \hat{\times} T_2$ is a barred tree.

3. Observe that if $a_1, a_2 \in \mathbb{N}^*$ have the same length then

$$\pi(a_1, a_2) \in T_1 \hat{\times} T_2 \;\Leftrightarrow\; \tau(a_1, a_2) \in T_1 \wedge T_2 \;\&\; (\pi_1 a_1 = \pi_1 a_2).$$

The result easily follows.

■

References

1. P. Aczel and M. Rathjen, Notes on Constructive Set Theory, Mittag-Leffler Technical Report No. 40, 2000/2001, (2001).
2. L.E.J. Brouwer, Über Definitionsbereiche von Funktionen, *Math. Ann. 95*, 453–472, (1927).
3. J. van Heijenoort, ed. From Frege to Gödel. A Sourcebook in Mathematical Logic 1879–1931, Harvard University Press, Cambridge Mass., Reprinted (1970).
4. N. Lusin, Sur les ensembles analytiques, *Fund. Math. 10*, 1–95, (1927).
5. N. Lusin, Leçons sur les ensembles analytiques et leurs applications, Collection de monographies sur la theorie des fonctions, Paris, Gauthiers-Villars (1930).
6. N. Lusin and W. Sierpinski, Sur quelques proprietés des ensembles (A), *Bull. Int. Acad. Sci. Cracovie, Série A; Sciences Mathématiques*, 35–48, (1918).
7. N. Lusin and W. Sierpinski, Sur un ensemble non mesurable B, *Journal de Mathématiques, 9^e^ serie, 2*, 53–72, (1923).
8. P. Martin-Löf, Notes on Constructive Mathematics, Almqvist & Wiksell, Stockholm, (1970).
9. Y. Moschovakis, Descriptive Set Theory, North Holland, Amsterdam, (1980).
10. M. Suslin, Sur une definition des ensembles mesurable B sans nombres transfinis, *Comptes Rendus Acad. Sci., Paris 164*, 88–91, (1917).
11. W. Veldman, Investigations in Intuitionistic Hierarchy Theory, Ph.D. thesis, Katholieke Universiteit, Nijmegen, (1981).
12. W. Veldman, The Borel Hierarchy and the Projective Hierarchy in Intuitionistic Mathematics, Report No. 0103, Department of Mathematics, University of Nijmegen, (March 2001), (revised version available, June 2006).

Dini's Theorem in the Light of Reverse Mathematics

Josef Berger and Peter Schuster

Abstract Dini's theorem says that compactness of the domain, a metric space, ensures the uniform convergence of every simply convergent monotone sequence of uniformly continuous real-valued functions whose limit is uniformly continuous. By showing that it is equivalent to Brouwer's fan theorem for detachable bars, we provide Dini's theorem with a classification in the constructive reverse mathematics recently propagated by Ishihara. If the functions occurring in Dini's theorem are pointwise continuous but integer-valued, then to still obtain such a classification we need to replace the fan theorem by the principle that every pointwise continuous integer-valued function on the Cantor space is uniformly continuous. As a complement, Dini's theorem both for pointwise and uniformly continuous functions is proved to be equivalent to the analogue of the fan theorem, weak König's lemma, in the classical setting of reverse mathematics started by Friedman and Simpson.

1 Introduction

Dini's theorem does not occur in the standard reference [19] for the programme of reverse mathematics founded by Friedman and Simpson. We now undertake a classification of Dini's theorem within the constructive reverse mathematics put forward by Ishihara [8–11].[1] In particular, we work in the constructive mathematics initiated by Bishop [3, 4, 6]. Bishop's theory can be seen as mathematics with intuitionistic logic in place of classical logic [16], in which vein 'classical' is sometimes used as a synonym for 'using the law of excluded middle'. Apart from the different choice of the underlying logic, one proceeds in Bishop's framework as in—the then dubbed classical—customary mathematics.

J. Berger (✉)
Mathematisches Institut, Universität München, Theresienstr. 39, 80333 München, Germany
e-mail: jberger,pschust@mathematik.uni-muenchen.de

This is a revised and extended version of our note [2]. As compared with its forerunner, the main feature of the present paper is the particular attention paid to the interplay between pointwise and uniform continuity.

[1] Related work has been done in parallel by Loeb [15] and Veldman [21].

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_7,

Our principal objective is to establish Dini's theorem for uniformly continuous real-valued functions as an equivalent of Brouwer's fan theorem for detachable bars. We first show that the latter is equivalent to the former for functions on the Cantor space. Only then we prove that the fan theorem is equivalent to Dini's theorem for uniformly continuous real-valued functions on every compact metric space or, alternatively, on the unit interval. Moreover, Dini's theorem for pointwise continuous but integer-valued functions turns out to be equivalent to the principle that every pointwise continuous integer-valued function on a compact metric space, or simply on the Cantor space, is uniformly continuous.

The implications which we assert from Remark 1 through Theorem 9, and Lemma 20 hold within $\mathrm{EL} + \mathrm{AC}_{00}$: that is, elementary analysis [20, Chapter 3, Section 6] enriched with number–number choice. For the sake of an easier reading, we do not encode finite sequences into integers, which is a routine task in the present context. We sometimes refer to work done directly in EL [10] or in a subsystem thereof [9], and to constructions from [6] and [20] which one can carry over to $\mathrm{EL} + \mathrm{AC}_{00}$ without any difficulty.

In addition to invoking unique and dependent choice,[2] unrestricted use is made in Bishop's setting of induction over the natural numbers. The latter principle distinguishes, among other things, Bishop's framework from the corresponding formal system RCA_0 in Simpson's classical hierarchy, in which induction is restricted to Σ_1^0-formulas [19, Remark I.8.9, Section I.12]. To conclude the present paper, we show with Theorem 21 that Dini's theorem both for pointwise and uniformly continuous functions is equivalent, in RCA_0, to the so-called weak König's lemma—the counterpart of Brouwer's fan theorem in classical reverse mathematics.

2 Some Alternative Formulations of Dini's Theorem

We adopt Bishop's definition of compactness [3, 4]: a metric space is said to be compact precisely when it is totally bounded and complete. In particular, every compact metric space is separable, and can thus be represented in terms of binary sequences [6, Chapter 5, Section 1][20, Chapter 7, Section 4]. On the other hand, we distinguish throughout between uniform and pointwise continuity even for mappings on compact metric spaces, rather than following Bishop's definition according to which a 'continuous' mapping on a compact metric space is uniformly continuous. Note that to give a uniformly continuous function on a compact metric space is the same as to give a uniformly continuous function on any dense subspace.

Throughout this note, let X be a metric space. We consider the following conclusion of Dini's theorem as a property of X:

[2] Richman [17, 18] has initiated a study of constructive mathematics without countable choice.

DT_X *If a monotone sequence (f_n) of uniformly continuous functions on X converges simply to a uniformly continuous function f on X, then (f_n) converges uniformly to f.*

So Dini's theorem says that if X is compact, then DT_X holds. Unless specified otherwise, all functions occurring in this paper are understood to be real-valued. If g and h are functions on X , write $g \leqslant h$ whenever $g(x) \leqslant h(x)$ for all $x \in X$, and likewise with $<$ in place of $\leqslant$.

One arrives at equivalents of DT_X if 'monotone' is replaced either by 'increasing' (that is, $f_n \leqslant f_{n+1}$ for all n) or by 'decreasing' (that is, $f_n \geqslant f_{n+1}$ for all n); one may further assume that $f = 0$. In particular, DT_X is equivalent to its following specific form:

For every decreasing sequence (g_n) of non-negative uniformly continuous functions on X, if (g_n) converges simply to 0, then (g_n) converges uniformly to 0.

The latter equivalent has the virtue that it allows for carrying over Dini's theorem to mappings on X with values in an arbitrary metric space Y. Given mappings $g, h : X \to Y$, define $d(g, h) : X \to \mathbb{R}$ by assigning $d(g(x), h(x))$ to every $x \in X$.

Remark 1 DT_X is equivalent to the validity of the following statement for all metric spaces Y: if (f_n) is a sequence of uniformly continuous mappings $f_n : X \to Y$ that converges simply to a uniformly continuous mapping $f : X \to Y$ such that $d(f_{n+1}, f) \leqslant d(f_n, f)$ for every $n \in \mathbb{N}$, then (f_n) converges uniformly to f.

Note that for being able to express DT_X in EL one needs to assume that the metric space X is complete and separable (and thus representable by a type 1 object), and that the functions f and f_n are all representable in type 1 objects (and thus automatically pointwise continuous).

3 Dini's Theorem for Functions on the Cantor Space

As usual, let $\{0, 1\}^{\mathbb{N}}$ denote the set of infinite binary sequences $\alpha, \beta, \ldots$, and let $\{0, 1\}^*$ stand for the set of finite binary sequences $u, v, w, \ldots$ The letters $k, \ell, m, M, n, N, p, q, \ldots$ are understood as variables ranging over the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of non-negative integers.

If $u \in \{0, 1\}^n$ for some n, then $|u| = n$ is the length of u. The n-th finite initial segment $\overline{\alpha}n = (\alpha(0), \ldots, \alpha(n-1))$ of α has length n, which includes the case $n = 0$ of the empty sequence. Concatenation of sequences is denoted by juxtaposition, and $w \geqslant u$ means that $w = uv$ for some v: that is, w is an extension of u and u is a restriction of w.

We know that $\{0, 1\}^{\mathbb{N}}$ is a compact metric space, the Cantor space, under the metric

$$d(\alpha, \beta) = \inf\{2^{-n} : \overline{\alpha}n = \overline{\beta}n\},$$

for which

$$d(\alpha, \beta) \leqslant 2^{-m} \iff \overline{\alpha}m = \overline{\beta}m \, .$$

So a function f on $\{0, 1\}^{\mathbb{N}}$ is uniformly continuous precisely when for every k there is m such that

$$\overline{\alpha}m = \overline{\beta}m \implies |f(\alpha) - f(\beta)| \leqslant 2^{-k}$$

for all α and β.

The open balls of $\{0, 1\}^{\mathbb{N}}$ are the subsets $\{\alpha : \alpha \in u\}$ with $u \in \{0, 1\}^*$, where $\alpha \in u$ is written in place of $\overline{\alpha}\,|u| = u$. Moreover, $\{0, 1\}^*$ is a countable dense subset of $\{0, 1\}^{\mathbb{N}}$, where every u is identified with its trivial extension $u00\ldots$. For a more detailed treatment of all this we refer to [6, Chapter 5, Section 3], [7, Section 3.2] , and [20, Chapter 4, Section 7].

To unwind Dini's theorem on the Cantor space, consider a decreasing sequence (g_n) of non-negative uniformly continuous functions on $\{0, 1\}^{\mathbb{N}}$. The implication

$$\textit{if } (g_n) \textit{ converges simply to } 0, \textit{ then } (g_n) \textit{ converges uniformly to } 0$$

crucial for the corresponding instance of $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ is equivalent to

$$\forall k \, \forall \alpha \, \exists n \, \left(g_n(\alpha) < 2^{-k}\right) \implies \forall k \, \exists N \, \forall \alpha \, \exists n \leqslant N \, \left(g_n(\alpha) < 2^{-k}\right). \tag{1}$$

(For we want to relate Dini's theorem to the fan theorem, we need to formulate this implication and their following equivalents in a way that at first glance may seem unnecessarily involved.) Furthermore, (1) is equivalent to

$$\forall k \, \forall \alpha \, \exists n \, \forall \beta \in \overline{\alpha}n \, \left(g_n(\beta) < 2^{-k}\right) \implies \forall k \, \exists N \, \forall \alpha \, \exists n \leqslant N \, \forall \beta \in \overline{\alpha}n \, \left(g_n(\beta) < 2^{-k}\right). \tag{2}$$

We now suppose that each g_n has a modulus of uniform continuity: that is, there is a sequence (M_{nk}) of non-negative integers with

$$\overline{\alpha}M_{nk} = \overline{\beta}M_{nk} \implies |g_n(\alpha) - g_n(\beta)| \leqslant 2^{-k} \tag{3}$$

for all α, β. By increasing the modulus if necessary, we can achieve that $M_{nk} \geqslant n$, and may thus set

$$W_k(u) = \left\{w : w \geqslant u \;\; \& \;\; |w| = M_{|u|k}\right\}$$

for every u and every k. For each $\beta \in u$ there is exactly one $w \in W_k(u)$ with $\beta \in w$, and we have $\left|g_{|u|}(\beta) - g_{|u|}(w)\right| \leqslant 2^{-k}$ for this $w = \overline{\beta}M_{|u|k}$. Hence (2) is equivalent to

$$\forall k \, \forall \alpha \, \exists n \, (\overline{\alpha}n \in U_k) \implies \forall k \, \exists N \, \forall \alpha \, \exists n \leqslant N \, (\overline{\alpha}n \in U_k) \tag{4}$$

with

$$U_k = \left\{u : \forall w \in W_k(u)\left(g_{|u|}(w) < 2^{-k}\right)\right\} .$$

Since $W_k(u)$ is a finite subset of $\{0, 1\}^*$, we have the simpler characterisation

$$U_k = \left\{u : G_k(u) < 2^{-k}\right\}$$

with

$$G_k(u) = \max\left\{g_{|u|}(w) : w \in W_k(u)\right\} .$$

As $\{0, 1\}^*$ is dense in $\{0, 1\}^{\mathbb{N}}$, we may further suppose that each g_n is given by a sequence $r_n = (r_{n\ell})$ of functions on $\{0, 1\}^*$ with rational values. The intended meaning is that for every u the sequence $r_n(u) = (r_{n\ell}(u))$ of rational numbers represents the real number $g_n(u)$ with $|g_n(u) - r_{n\ell}(u)| < 2^{-\ell}$ for all ℓ. We thus require the presence of a sequence of functions $r_{n\ell} : \{0, 1\}^* \to \mathbb{Q}$ with

$$\left|r_{np}(u) - r_{nq}(u)\right| < 2^{-p} + 2^{-q} . \tag{5}$$

The conditions $g_n \geqslant 0$ and $g_{n+1} \geqslant g_n$ can then be put as

$$r_{n\ell}(u) \geqslant -2^{-\ell} \quad \text{and} \quad r_{n+1,2\ell}(u) \geqslant r_{n,2\ell}(u) - 2^{-\ell}, \tag{6}$$

respectively. To express that each g_n is uniformly continuous, we may assume that the modulus of uniform continuity (M_{nk}) with $M_{nk} \geqslant n$ from (3) works also for r_n: that is,

$$\overline{u}M_{nk} = \overline{v}M_{nk} \implies \forall \ell \left(\left|r_{n,2\ell}(u) - r_{n,2\ell}(v)\right| < 2^{-k} + 2^{-\ell}\right) . \tag{7}$$

A decreasing sequence (g_n) of non-negative uniformly continuous functions on $\{0, 1\}^{\mathbb{N}}$ can therefore be identified with

(*) a sequence of functions $r_{n\ell} : \{0, 1\}^* \to \mathbb{Q}$ and a sequence of integers M_{nk} for which $M_{nk} \geqslant n$ and which satisfy the conditions (5), (6), and (7).[3]

In particular, the real number $G_k(u)$ is given by the sequence $(R_{k\ell}(u))$ of rational numbers

$$R_{k\ell}(u) = \max\left\{r_{|u|\ell}(w) : w \in W_k(u)\right\} .$$

[3] Needless to say, these sequences can be put as functions $r : \mathbb{N} \times \mathbb{N} \times \{0, 1\}^* \to \mathbb{Q}$ and $M : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ with $r(n, \ell, u) = r_{n\ell}(u)$ and $M(n, k) = M_{nk}$ —or, by means of an appropriate coding of finite sequences, as functions of type $\mathbb{N} \to \mathbb{N}$.

Hence $G_k(u) < 2^{-k}$ means that $R_{k\ell}(u) + 2^{-\ell} < 2^{-k}$ for some ℓ, and $u \in U_k$ corresponds to $u \in A_k$ with

$$A_k = \left\{u : \exists \ell \ \left(R_{k\ell}(u) + 2^{-\ell} < 2^{-k}\right)\right\} .$$

In all, to assert $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ means that all data of type (*) satisfy

$$\forall k \, \forall \alpha \, \exists n \ (\overline{\alpha}n \in A_k) \implies \forall k \, \exists N \, \forall \alpha \, \exists n \leqslant N \ (\overline{\alpha}n \in A_k) . \tag{8}$$

Note that (8) is the counterpart of (4).

4 The Fan Theorem as an Equivalent of Dini's Theorem

A subset B of $\{0, 1\}^*$ is detachable if $u \in B$ is a decidable predicate of $u \in \{0, 1\}^*$: that is, for each u either $u \in B$ or else $u \notin B$. To give a detachable subset B of $\{0, 1\}^*$ is the same as to give its characteristic function $\chi_B : \{0, 1\}^* \to \{0, 1\}$ with $\chi_B(u) = 1$ precisely when $u \in B$.

Moreover, a subset B of $\{0, 1\}^*$ is a bar if for every α there is n with $\overline{\alpha}n \in B$, while a bar B is uniform if there exists N such that for every α there is $n \leqslant N$ with $\overline{\alpha}n \in B$. Brouwer's fan theorem for detachable bars reads as follows:

FT *Every detachable bar is uniform.*

Another way to put FT is to require the validity of the implication

$$\forall \alpha \, \exists n \ (\overline{\alpha}n \in B) \implies \exists N \, \forall \alpha \, \exists n \leqslant N \ (\overline{\alpha}n \in B) \tag{9}$$

from all detachable subsets B of $\{0, 1\}^*$. We refer to [10] for formal versions of the notion of a detachable subset of $\{0, 1\}^*$ and of the more specific notion which occurs next.

One arrives at an equivalent of FT by restricting it to the subsets B of $\{0, 1\}^*$ which are closed under extension: that is, if $u \in B$ and $w \geqslant u$, then $w \in B$. Every B satisfying this extra condition is a uniform bar precisely when there exists N such that $\overline{\alpha}N \in B$ for every α.

Lemma 2 [10, Lemma 1] FT *is equivalent to the statement that the implication*

$$\forall \alpha \, \exists n \ (\overline{\alpha}n \in B) \implies \exists N \forall \alpha \ (\overline{\alpha}N \in B) \tag{10}$$

holds for all detachable subsets B of $\{0, 1\}^*$ *which are closed under extension.*

Note the difference between (10) and (9).

Proposition 3 FT *follows from* $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$.

Proof We use Lemma 2. Let B be a detachable subset of $\{0, 1\}^*$ that is closed under extension, and assume that B is a bar. For every n define $f_n : \{0, 1\}^{\mathbb{N}} \to$

$\{0, 1\}$ by setting $f_n(\alpha) = \chi_B(\overline{\alpha}n)$. Each f_n is uniformly continuous, because $f_n(\alpha)$ depends—for n fixed—only on $\overline{\alpha}n$. In addition, the sequence (f_n) is increasing and converges simply to 1. Hence the convergence is uniform, which is to say that B is a uniform bar. Q.E.D.

To show the reverse implication, FT needs to be extended to the subsets B of $\{0, 1\}^*$ which—in the terminology of [6]—are simply existential: that is, there is a sequence (C_ℓ) of detachable subsets of $\{0, 1\}^*$ such that $u \in B$ precisely when $u \in C_\ell$ for some ℓ. This condition is equivalent to the existence of a detachable subset C of $\{0, 1\}^* \times \mathbb{N}$ such that $u \in B$ if and only if $(u, \ell) \in C$ for some ℓ.

The following is contained in [9, Proposition 16.15]. We do a proof without coding.

Lemma 4 FT *is equivalent to the statement that the implication*

$$\forall\alpha\, \exists n\ (\overline{\alpha}n \in B) \implies \exists N\, \forall\alpha\, \exists n \leqslant N\ (\overline{\alpha}n \in B) \tag{11}$$

holds for all simply existential subsets B of $\{0, 1\}^*$.

Proof Only one direction needs a proof. Let B be a simply existential subset of $\{0, 1\}^*$, and pick a detachable subset C of $\{0, 1\}^* \times \mathbb{N}$ for which $u \in B$ if and only if $(u, \ell) \in C$ for some ℓ. Set

$$D = \{u : \exists n, \ell \leqslant |u|\, ((\overline{u}n, \ell) \in C)\},$$

which is a detachable subset of $\{0, 1\}^*$. If B is a bar, then so is D (because if $(\overline{\alpha}n, \ell) \in C$, then $\overline{\alpha}m \in D$ with $m = \max\{n, \ell\}$). On the other hand, if D is a uniform bar, then so is B (because if $\overline{\alpha}m \in D$ for some $m \leqslant N$, then $(\overline{\alpha}n, \ell) \in C$ for certain $n, \ell \leqslant m \leqslant N$). So if (9) holds with D in place of B, then (11) follows. Q.E.D.

As (11) and (9) are identical, FT can also be put as 'every simply existential bar is uniform'.

Proposition 5 FT *implies* $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$.

Proof Assume that we are given data of type (*). To arrive at (8), it suffices to achieve

$$\forall\alpha\, \exists n\ (\overline{\alpha}n \in A_k) \implies \exists N\, \forall\alpha\, \exists n \leqslant N\ (\overline{\alpha}n \in A_k) \tag{12}$$

for arbitrary but fixed k. Since $R_{k\ell}(u)$ is a rational number, the condition $R_{k\ell}(u) + 2^{-\ell} < 2^{-k}$ is a decidable property of u for any given ℓ. Hence A_k is a simply existential subset of $\{0, 1\}^*$, and Lemma 4 applies. Q.E.D.

Corollary 6 FT *and* $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ *are equivalent.*

Since FT already follows from $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ for functions with values in $\{0, 1\}$ (see the proof of Proposition 3), Corollary 6 would still hold if one restricted $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ to functions with values in $\{0, 1\}$ or, more generally, in $\mathbb{N}$. The equivalence of FT and $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ for functions with values in $\mathbb{N}$ has also been shown by Veldman [21].

Proposition 7 $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ *entails* DT_X *for all compact metric spaces* X.

Proof Every compact metric space X is a uniformly continuous image of the Cantor space [6, Chapter 5, Theorem (1.4)] [20, Chapter 7, Corollary 4.4]: that is, there is a uniformly continuous mapping from $\{0, 1\}^{\mathbb{N}}$ onto X. Along any such mapping, DT_X can be deduced from $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ in the obvious way. Q.E.D.

Proposition 8 $\mathrm{DT}_{[0,1]}$ *implies* FT.

Proof Let B be a bar. By [12] (see also [6, Chapter 6, Theorem (2.7)]) there is a uniformly continuous function h on $[0, 1]$ with $h > 0$ such that $\inf h > 0$ if and only if B is uniform. Set $f = 1 - \min\{h, 1/2\}$, for which $0 < f < 1$. Since $\inf h > 0$ if and only if $\sup f < 1$, it suffices to show that f is bounded away from 1. To this end, set $f_n = 1 - f^n$ for every n. Note that f and all the f_n are uniformly continuous, and that (f_n) is increasing and converges simply to 1. By hypothesis, (f_n) converges uniformly to 1; whence $f^n = 1 - f_n < 1/2$ and thus $f < \sqrt[n]{1/2}$ for some n. Q.E.D.

The idea underlying the foregoing proof stems from the recursive counterexample to Dini's theorem that Bridges [5] ascribes to Richman.

To simplify the presentation, we define the following conjunction:

DT *Every compact metric space* X *satisfies* DT_X.

Theorem 9 *The following four items are equivalent:* FT; $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$; DT; $\mathrm{DT}_{[0,1]}$.

5 Dini's Theorem and Pointwise Versus Uniform Continuity

Working in an informal setting, we now consider the conclusion of Dini's theorem for pointwise continuous functions as a property of any metric space X:

DT_X^+ *If a monotone sequence* (f_n) *of pointwise continuous functions on* X *converges simply to a pointwise continuous function* f *on* X, *then* (f_n) *converges uniformly to* f.

Likewise, one is prompted to look at the uniform continuity principle as a property of X:

UC_X *Every pointwise continuous function on* X *is uniformly continuous.*

While DT_X^+ implies DT_X for every X, the converse implication is valid under UC_X.

Lemma 10 *If* UC_X *holds for a metric space* X, *then* DT_X^+ *and* DT_X *are equivalent.*

A function f on X is bounded if $|f| \leqslant N$ for some N, where $|f|$ stands for the function assigning $|f(x)|$ to every $x \in X$. We need to consider yet another principle as a property of any metric space X:

PB$_X$ *Every pointwise continuous function on X is bounded.*

Proposition 11 DT_X^+ *entails* PB_X *for every metric space X.*

Proof If f is a pointwise continuous function on X, then so is $f_n = \min\{|f|, n\}$ for every n. The sequence (f_n) is increasing and converges simply to $|f|$. By DT_X^+, this convergence is uniform; whence there is $N \in \mathbb{N}$ such that $|f| - 1 \leqslant f_N$ and thus $|f| \leqslant N + 1$. Q.E.D.

Proposition 12 *If X is totally bounded, then* UC_X *implies* PB_X.

Proof Uniformly continuous mappings preserve total boundedness [6, Chapter 2, Proposition (4.3)]. Q.E.D.

A function $f : \{0, 1\}^{\mathbb{N}} \to \mathbb{N}$ is pointwise continuous if and only if

$$\forall\alpha\exists n\forall\beta \in \overline{\alpha}n\ (f(\beta) = f(\overline{\alpha}n))\ ,$$

and uniformly continuous if and only if

$$\exists N\forall\alpha\forall\beta \in \overline{\alpha}N\ (f(\beta) = f(\overline{\alpha}N))\ .$$

Proposition 13 FT *follows from* $\mathrm{PB}_{\{0,1\}^{\mathbb{N}}}$ *for functions with values in* $\mathbb{N}$.

Proof If B is a detachable bar, then the function

$$f : \{0, 1\}^{\mathbb{N}} \to \mathbb{N}\,,\ \alpha \mapsto \min\{n : \overline{\alpha}n \in B\}$$

is well-defined. Since f is pointwise continuous, by $\mathrm{PB}_{\{0,1\}^{\mathbb{N}}}$ it is bounded; whence B is a uniform bar. Q.E.D.

Proposition 14 $\mathrm{UC}_{\{0,1\}^{\mathbb{N}}}$ *implies* $\mathrm{DT}^+_{\{0,1\}^{\mathbb{N}}}$.

Proof According to Lemma 10, it suffices to show that $\mathrm{UC}_{\{0,1\}^{\mathbb{N}}}$ implies $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$. If $\mathrm{UC}_{\{0,1\}^{\mathbb{N}}}$ holds, then so does FT (Propositions 12 and 13), from which $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ follows (Proposition 5). Q.E.D.

As we have done before in the case of **DT**, we consider the following conjunctions:

DT$^+$ *Every compact metric space X satisfies* DT_X^+.

UC *Every compact metric space X satisfies* UC_X.

PB *Every compact metric space X satisfies* PB_X.

Proposition 15 DT^+ *is equivalent to* $\mathrm{DT}^+_{\{0,1\}^{\mathbb{N}}}$, UC *is equivalent to* $\mathrm{UC}_{\{0,1\}^{\mathbb{N}}}$*, and* PB *is equivalent to* $\mathrm{PB}_{\{0,1\}^{\mathbb{N}}}$.

Proof The cases of DT^+ and PB can be treated as the one of DT in the proof of Proposition 7. A closer look at the references given therein [6, Chapter 5, Theorem (1.4)] [20, Chapter 7, Corollary 4.4] suffices for deducing UC from $\mathrm{UC}_{\{0,1\}^{\mathbb{N}}}$. In fact, for every compact metric space X there is a uniform quotient map π from $\{0,1\}^{\mathbb{N}}$ onto X, which is uniformly continuous and has the property that a function f on X is uniformly continuous already if $f \circ \pi$ is uniformly continuous. Q.E.D.

The following principle says that one can decide whether any given pointwise continuous $f : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ is constant:

DC *For every pointwise continuous function* $g : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ *either* $g = 0$ *or else* $\neg (g = 0)$.

In fact, if $f : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ is pointwise continuous, then f is constant precisely when $g = f - f(0)$ vanishes identically.

Proposition 16 DC *follows from* $\mathrm{PB}_{\{0,1\}^{\mathbb{N}}}$ *for functions with values in* $\mathbb{N}$.

Proof If $g : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ is pointwise continuous, then the function

$$h : \{0,1\}^{\mathbb{N}} \to \mathbb{N}\,,\ \alpha \mapsto \max\left(\{0\} \cup \{n : g(\overline{\alpha}n) \neq g(\alpha)\}\right)$$

is well-defined. Since h is pointwise continuous, by $\mathrm{PB}_{\{0,1\}^{\mathbb{N}}}$ it is bounded. To know whether $g = 0$, it therefore suffices to test for finitely many n whether $g(\overline{\alpha}n) = 0$. Q.E.D.

Corollary 17 *Each of the following four items implies the next:* UC*;* DT^+*;* PB*;* FT + DC.

To achieve the converse implications, we need to focus on integer-valued functions.

Proposition 18 FT + DC *proves* $\mathrm{UC}_{\{0,1\}^{\mathbb{N}}}$ *for functions with values in* $\mathbb{N}$.

Proof If $f : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ is any function, then

$$B = \{u \in \{0,1\}^* : \forall \beta \in u\ (f(\beta) = f(u))\}$$

is closed under extension. Moreover, B is a bar if and only if f is pointwise continuous, and B is a uniform bar if and only if f is uniformly continuous. If now f is pointwise continuous, then so is

$$f|_u : \{0,1\}^{\mathbb{N}} \to \mathbb{N}\,,\ \gamma \mapsto f(u\gamma)$$

for every $u \in \{0,1\}^*$, and $f|_u$ is constant precisely when $\forall \beta \in u\ (f(\beta) = f(u))$ holds. Hence B is, by DC, a detachable bar and thus, by FT, a uniform bar, which is to say that f is uniformly continuous. Q.E.D.

Theorem 19 *The following four items are equivalent:*

(i) UC *for functions with values in* $\mathbb{N}$*;*
(ii) DT^+ *for functions with values in* $\mathbb{N}$*;*
(iii) PB *for functions with values in* $\mathbb{N}$*;*
(iv) FT + DC.

The equivalence of (i), (iii), and (iv) has already been shown by the first author [1].

6 The Relation of Dini's Theorem to Weak König's Lemma

A (binary) tree is a detachable subset T of $\{0, 1\}^*$ which contains the empty sequence, and which is closed under restriction: that is, if $w \in T$ and $u \leqslant w$, then $u \in T$. A tree T is infinite if for every n there is $u \in T$ with $|u| = n$, and an infinite path of T is an α with $\overline{\alpha}n \in T$ for all n.

Weak König's lemma in Simpson's terminology [19] is the following statement:

WKL *Every infinite tree has an infinite path.*

A tree T is infinite precisely when for every n there is α with $\overline{\alpha}n \in T$. Hence to postulate WKL amounts to require the validity of the implication

$$\forall n\, \exists \alpha\ (\overline{\alpha}n \in T) \implies \exists \alpha\, \forall n\ (\overline{\alpha}n \in T) \tag{13}$$

from all trees T.

It is known that WKL and FT are the classical contrapositives of each other. Ishihara has even proved that WKL implies FT over EL [10]. For the sake of completeness, we shed some light on the classical equivalence of WKL and FT, following [10].

A tree T is finite if there is N such that $|u| < N$ for every $u \in T$, and without infinite path if for every α there is n with $\overline{\alpha}n \notin T$. These are classically equivalent ways to express that a tree is not infinite and has no infinite path, respectively. The classical contrapositive of WKL can therefore be put as follows:

WKL$^\neg$ *Every tree without infinite path is finite.*

A tree T is finite if and only if there is N such that $\overline{\alpha}N \notin T$ for all α. To assert WKL$^\neg$ thus amounts to require the validity of the implication

$$\forall \alpha\, \exists n\ (\overline{\alpha}n \notin T) \implies \exists N\, \forall \alpha\ (\overline{\alpha}N \notin T) \tag{14}$$

from all trees T. Note that (13) and (14) are the classical contrapositives of each other.

If $\{0, 1\}^*$ is the disjoint union of two inhabited subsets B and T, then B is closed under extension precisely when T is closed under restriction (that is, T is a tree), B is a bar if and only if T has no infinite path, and B is a uniform bar if and only if T is a finite tree. In other words, (14) is the same as (10) for any such choice of B and T; whence Lemma 2 can be rephrased as follows.

Lemma 20 $\mathrm{WKL}^{\neg}$ *and* FT *are equivalent.*

During the rest of this paper we work within the formal system RCA_0 from [19], whose notations and conventions we adopt. In particular, we switch from constructive to classical reverse mathematics, and restrict our attention to pointwise continuous functions that are representable. Note that the latter move slightly changes the meaning of DT^+ and UC. Also, AC_{00} is not admissible in RCA_0, since otherwise WKL would be equivalent to a fragment of the law of excluded middle [9, Proposition 16.18, Theorem 16.21] and thus be valid in RCA_0 as a system based on classical logic. Unlike EL, moreover, the system RCA_0 only disposes of restricted induction. All these aspects will be taken into account when proving the next result.

Theorem 21 *In* RCA_0*, the following three items are equivalent:* DT*;* DT^+*;* WKL.

Proof By combining Lemma 20 with Proposition 3, in whose proofs we have used neither unrestricted induction nor AC_{00}, one can deduce WKL from $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ in RCA_0. According to [19, Theorem IV.2.2], WKL proves UC in RCA_0; whence it suffices (see Lemma 10) to deduce DT from WKL in RCA_0. To this end, we mimic the proof of [19, Theorem IV.2.2] as follows.

Let $X = \widehat{A}$ be a compact metric space. Suppose that (g_n) is a decreasing sequence of non-negative uniformly continuous functions on X which converges simply to 0. Let $\varphi(n, a, r, m)$ be a Σ_1^0-formula which says that $a \in A$, $r \in \mathbb{Q}^+$, $m \in \mathbb{N}$, and that

> (§) there are $b \in \mathbb{Q}$ and $s \in \mathbb{Q}^+$ with $b < 2^{-n-1}$ and $s < 2^{-n-1}$ such that $(a, r)\, g_m\, (b, s)$.

One can show that

> (†) for every $x \in X$ and every n there are a, r, m with $\varphi(n, a, r, m)$ and $d(x, a) < r$.

By [19, Lemma II.3.7], there is a sequence $\langle (a_{ni}, r_{ni}, m_{ni}) : i, n \in \mathbb{N} \rangle$ such that $\varphi(n, a, r, m)$ if and only if $(a, r, m) = (a_{ni}, r_{ni}, m_{ni})$ for some i. By (†), $\langle \langle B(a_{ni}, r_{ni}) : i \in \mathbb{N} \rangle : n \in \mathbb{N} \rangle$ is a sequence of open coverings of X, which—according to [19, Theorem IV.1.6]—gives rise to a sequence of finite subcoverings $\langle \langle B(a_{ni}, r_{ni}) : i \leqslant k_n \rangle : n \in \mathbb{N} \rangle$. If we now set $N_n = \max\{m_{ni} : i \leqslant k_n\}$ for every n, then $g_{N_n}(x) < 2^{-n}$ for every $x \in X$. In fact, for every n and every x there is $i \leqslant k_n$ with $x \in B(a_{ni}, r_{ni})$, for which $g_{m_{ni}}(x)$ belongs to the closure of $B(b, s)$ for some b and s as in (§) with $(n, a_{ni}, r_{ni}, m_{ni})$ in place of (n, a, r, m). Since, in particular, $b < 2^{-n-1}$ and $s < 2^{-n-1}$, we have $g_{N_n}(x) \leqslant g_{m_{ni}}(x) \leqslant b + s < 2^{-n}$ as required. In other words, (g_n) converges uniformly to 0. Q.E.D.

Kohlenbach [14] deduces DT_X for $X = [0, 1]^n$ with $n \in \mathbb{N}$ from a strong principle of uniform boundedness, which he extracts from a generalisation of WKL to higher types [13].

7 Discussion

Bishop's concept of a (uniformly) continuous function on a compact metric space includes a modulus of uniform continuity, whose existence is guaranteed anyhow in the presence of countable choice for natural numbers. On the other hand, 'it is interesting to note that "any continuous function which arises in practice" can be proved in RCA_0 to have a modulus', while 'in general its existence is not provable in RCA_0' [19, Remark IV.2.8]. In the present paper, the additional information given by a modulus was only needed on our way from FT to $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$, whereas no modulus occurred at any other place—let alone during the corresponding argument, which we gave in RCA_0, that DT follows from WKL.

We anyway hold Brouwer's fan theorem for conceptually more appropriate than weak König's lemma to classify uniformity theorems such as Dini's. While FT provides us with a uniform bound—a single natural number—in a way similar to $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$, the conclusion of WKL consists of the existence of an object of a different nature: an infinite sequence. More specifically, the logical form (8) of $\mathrm{DT}_{\{0,1\}^{\mathbb{N}}}$ corresponds rather to (9) than to (13). One cannot make this distinction unless one moves from classical to constructive reverse mathematics.

Acknowledgments During the preparation of [2] the first author was supported by a grant of the Graduiertenkolleg *Logik in der Informatik* hosted by the *Deutsche Forschungsgemeinschaft.* The way to a smooth classification of Dini's theorem was paved by Ulrich Kohlenbach's suggestion to focus first on uniform continuity. The present work further benefited from Hajime Ishihara's repeated advice in issues of formalisation, and from discussions with Thierry Coquand, Antonino Drago, Martin Hyland, Joan R. Moschovakis, Thomas Streicher, and Rudolf Taschner. Particular credit is given to the anonymous referees of [2] and to the one of the present paper for their considerable help with bringing both items into the final form.

References

1. Berger, J., Constructive equivalents of the uniform continuity theorem. *J. UCS* 11, 1878–1883, 2005
2. Berger, J., and P. Schuster, Classifying Dini's theorem. *Notre Dame J. Formal Logic* 47, 253–262, 2006
3. Bishop, E., *Foundations of Constructive Analysis*. McGraw–Hill, New York, 1967
4. Bishop, E., and D. Bridges, *Constructive Analysis*. Springer, Berlin etc., 1985
5. Bridges, D.S., Dini's theorem: a constructive case study. In: C.S. Calude et al., eds., *Combinatorics, Computability and Logic*. 3rd International Conference DMTCS01, Constanţa, Romania, 2001. Proceedings. Springer, London, 69–80, 2001
6. Bridges, D., and F. Richman, *Varieties of Constructive Mathematics*. Cambridge University Press, Cambridge, 1987

7. Dummett, M., *Elements of Intuitionism.* 2nd ed., Oxford University Press, Oxford, 2000
8. Ishihara, H., Informal constructive reverse mathematics. *Sūrikaisekikenkyūsho Kōkyūroku* 1381, 108–117, 2004
9. Ishihara, H., Constructive reverse mathematics: compactness properties. In: L. Crosilla et al., eds., *From Sets and Types to Topology and Analysis.* Oxford University Press, Oxford, 245–267, 2005
10. Ishihara, H., Weak König's lemma implies Brouwer's fan theorem: a direct proof. *Notre Dame J. Formal Logic* 47, 249–252, 2006
11. Ishihara, H., Reverse mathematics in Bishop's constructive mathematics. *Philosophia Scientiae*, cahier special 6 (2006), 43–59
12. Julian, W.H., and F. Richman, A uniformly continuous function on [0, 1] which is everywhere different from its infimum. *Pacific J. Math* 111, 333–340, 1984
13. Kohlenbach, U., Mathematically strong subsystems of analysis with low rate of growth of provably recursive functionals. *Arch. Math. Logic.* 36, 31–71, 1996
14. Kohlenbach, U., The use of a logical principle of uniform boundedness in analysis. In: A. Cantini et al., eds., *Logic and Foundations of Mathematics.* Kluwer, Dordrecht, 93–106, 1999
15. Loeb, I., Equivalents of the (weak) fan theorem. *Ann. Pure Appl. Logic* 132, 51–66, 2005
16. Richman, F., Intuitionism as generalization. *Philos. Math.* 5, 124–128, 1990
17. Richman, F., The fundamental theorem of algebra: a constructive development without choice. *Pacific J. Math.* 196, 213–230, 2000
18. Richman, F., Constructive mathematics without choice. In: P. Schuster et al., eds., *Reuniting the Antipodes. Constructive and Nonstandard Views of the Continuum.* Kluwer, Dordrecht, 199–205, 2001
19. Simpson, S.G., *Subsystems of Second Order Arithmetic*. Springer, Berlin etc., 1999
20. Troelstra, A.S., and D. van Dalen, *Constructivism in Mathematics*. Two volumes. North-Holland, Amsterdam, 1988
21. Veldman, W., Brouwer's fan theorem as an axiom and as a contrast to Kleene's alternative. Preprint, Radboud University, Nijmegen, 2005

Journey into Apartness Space

Douglas Bridges and Luminiţa Simona Vîţă

Abstract We present some of the fundamental notions and results in the axiomatic theory of apartness spaces, a constructive approach to topology. The paper begins with apartness between sets, and between points and sets, and ends with very recent work on the theory of apartness on frames.

1 Introduction

We begin with two quotes from Errett Bishop (1928–84):

> Very little is left of general topology after that vehicle of classical mathematics has been taken apart and reassembled constructively. With some regret, plus a large measure of relief, we see this flamboyant engine collapse to constructive size. ([2], p. 63)
> The problem of finding a suitable constructive framework for general topology is important and elusive. (Letter to Bridges, 14 April 1975)

Perhaps as a result of the first of these statements, together with Bishop's suggestion, in Appendix A to [2], that constructive theories like that of distributions might rely on *ad hoc* topological notions, little constructive attention was paid to general topology over the period 1967–2000. (There were, however, notable exceptions, such as the work of Grayson [13, 14], the "point-free" approach adopted by Sambin, Martin-Löf and others [19, 20], and Waaldijk's thesis [24].)

In the mid–1970s, Bridges came across an article advocating the use of nearness in the teaching of a first course in analysis [10]. This led him to consider nearness, or rather the constructively more appropriate opposite notion of apartness, as a possible foundation for constructive topology. Not, at that stage, being sufficiently mathematically experienced to make significant progress with this idea, he put it on one side until February 2000, when he and Vîţă began the project on apartness spaces that is discussed in the present exposition. Clearly, two heads were better than one: for in the intervening eight years, the theory has developed substantially,

D. Bridges (✉)
Department of Mathematics & Statistics, University of Canterbury,
Private Bag 4800, Christchurch, New Zealand
e-mail: d.bridges@math.canterbury.ac.nz

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_8,

with contributions from Hajime Ishihara, Peter Schuster, Fred Richman, and several others.[1]

In this paper we outline some of the main ideas of the theories of apartness between points and sets, sets and sets, and (the most recent development) elements of a lattice. We assume that the reader understands the basic differences between constructive and classical mathematics, and that the former is essentially mathematics carried out with intuitionistic logic and some appropriate set theory such as the Aczel-Rathjen CST [1].[2] In practice, the paper can be read without any deep knowledge of either constructive analysis or classical topology, although we imagine that anyone interested in reading it will know plenty about the latter.

2 Set–Set Pre-Apartness

Although the development of apartness theory began with point–set apartness, we shall adopt an unhistorical approach in which (pre-)apartness between sets is the primary notion, point–set apartness being a special case thereof.

A **pre-apartness space** is an inhabited set X with an inequality and a symmetric[3] binary relation $\bowtie$ between subsets of X that satisfies the following four axioms.

B1 $X \bowtie \varnothing$.
B2 $A \bowtie B \Rightarrow A \subset {\sim}B$
B3 $A \bowtie (B \cup C) \Longleftrightarrow A \bowtie B \wedge A \bowtie C$
B4 $-A \subset {\sim}B \Rightarrow -A \subset -B$,

where

$$\sim S = \{x \in X : \forall y \in S\,(x \neq y)\}$$

is the **complement**, and

$$-S = \{x \in X : \{x\} \bowtie S\} \tag{1}$$

the **apartness complement**, of S. The relation $\bowtie$ is called the **pre-apartness** on X.

Defining

$$x \bowtie A \Longleftrightarrow \{x\} \bowtie A, \tag{2}$$

[1] Peter Schuster recently brought to our attention the paper [15], in which some of our axioms of apartness are prefigured.

[2] Our development of the theory of apartness spaces might appear to require the power-set axiom, which, on the grounds of its impredicativity, is excluded from CST. However, we believe that, with suitable modifications at the formal level, our theory could be presented entirely within CST.

[3] The requirement of symmetry is dropped in our forthcoming book [7]. In order to simplify the exposition, we retain symmetry in this paper except when we deal with apartness on lattices.

we obtain a relation between points and subsets of X that satisfies the following properties of a **point–set pre-apartness**:

A1 $x \bowtie \varnothing$
A2 $-A \subset {\sim}A$
A3 $x \bowtie (A \cup B) \Longleftrightarrow x \bowtie A \wedge x \bowtie B$
A4 $-A \subset {\sim}B \Rightarrow -A \subset -B$.

There are many simple consequences of the axioms **B1–B4** and **A1**–**A4** that we shall use without further comment.

The canonical example of a set–set pre-apartness space is a metric space (X, ρ), where the apartness between subsets A, B is defined by

$$A \bowtie B \Longleftrightarrow \exists r > 0 \forall x \in A \forall y \in B \,(\rho(x, y) \geqslant r)\,.$$

We shall generalise this to uniform spaces later. Another example is provided by a topological space[4] (X, τ) taken with the set–set operation defined by

$$A \bowtie B \Longleftrightarrow (A \subset ({\sim}B)^{\circ} \vee B \subset ({\sim}A)^{\circ})\,. \tag{3}$$

Classically, this pre-apartness satisfies

$$A \bowtie B \Longleftrightarrow \left(A \cap \overline{B} = \varnothing \vee \overline{A} \cap B = \varnothing\right).$$

The condition on the right-hand side of this equivalence is constructively weaker than its counterpart in (3).

By a (set–set) **apartness** on X we mean a relation $\bowtie$ between subsets of X that satisfies the axioms **B1**–**B3** and

B5 $x \in -A \Rightarrow \exists S \subset X (x \in -S \wedge X = -A \cup S)$.

It then also satisfies **B4** and so is a pre–apartness. Metric spaces satisfy **B5**. A point–set pre-apartness that satisfies a weaker property than **B5**, namely

A5 $x \in -A \Rightarrow \forall y \in X\,(x \neq y \vee y \in -A)$,

is called a **point–set apartness**. A point–set apartness that satisfies **B5** is said to be **locally decomposable**.

An example in [3] uses the Smirnov topology to show that axiom **A5** is independent of **A1–A4**. Pre-apartness spaces that do not satisfy **A5** arise in practice: the real line with $\sim$ taken as the denial of equality is one. It is therefore appropriate to take note of those proofs that do not depend on **A5**.

The following strong form of axiom **B4/A4**,

B4$_s$ $(A \bowtie B \wedge -B \subset {\sim}C) \Rightarrow A \bowtie C$,

---

[4] We assume that every topological space comes equipped with an inequality relation.

holds in many important pre-apartness spaces. We illustrate the application of the axioms by considering **B4**$_s$ further, in a series of small results in a pre-apartness space X.

Lemma 1 $-S = -{\sim} - S$.

Proof Since $-S \subset {\sim}S$, we have $-{\sim} - S \subset -{\sim}{\sim}S$. Moreover, $S \subset {\sim}{\sim}S$, so $-{\sim}{\sim}S \subset -S$. Hence $-{\sim} - S \subset -S$. On the other hand, $-S \subset {\sim}({\sim} - S)$, so, by **B4**, $-S \subset -{\sim} - S$.

Lemma 2 ***B4***$_s$ *holds if and only if, for all* $A, B \subset X$,

$$A \bowtie B \Rightarrow A \bowtie {\sim} - B. \tag{4}$$

Proof If **B4**$_s$ holds and $A \bowtie B$, then since, by Lemma 1,

$$-B = -{\sim} - B \subset {\sim}({\sim} - B),$$

we see that $A \bowtie {\sim} - B$. Conversely, suppose that (4) holds, and let $A \bowtie B$ and $-B \subset {\sim}C$. Then $C \subset {\sim}{\sim}C \subset {\sim} - B$ and therefore (by **B3**), $A \bowtie C$.

Lemma 3 *Suppose that* **A5** *holds. Then*

$$\forall A, B, C \subset X\,(-A \subset \neg B \Rightarrow -A \subset {\sim}B). \tag{5}$$

Proof If $x \in -A$ and $y \in B$, then, by **A5**, either $x \neq y$ or $y \in -A \subset \neg B$; since the latter is absurd, it follows that $x \neq y$. Hence $-A \subset {\sim}B$. ■

Lemma 4 *Suppose that* **A5** *holds. Then* ***B4***$_s$ *is equivalent to the statement*

$$\forall A, B, C \subset X\,((A \bowtie B \wedge -B \subset \neg C) \Rightarrow A \bowtie C). \tag{6}$$

Proof Since ${\sim}C \subset \neg C$, it is clear that (6) implies **B4**$_s$ (even without the assumption of **A5**). The converse follows from Lemma 3. ■

Proposition 5 *If* B4$_s$ *holds, then*

$$\forall A, B \subset X\,(A \bowtie B \Rightarrow A \bowtie {\sim}{\sim}B).$$

If also A5 *holds, then*

$$\forall A, B \subset X\,(A \bowtie B \Rightarrow A \bowtie \neg\neg B).$$

Proof The first part of the proposition follows from the inclusion ${\sim}{\sim}B \subset {\sim} - B$ and Lemma 2. Now assume also **A5**. If $A \bowtie B$, then since $-B \subset \neg B = \neg(\neg\neg B)$, it follows from Lemma 4 that $A \bowtie \neg\neg B$. ■

The strongest of all the separation properties normally considered for an apartness space X is the **Efremovič condition:**

EF $A \bowtie B \Rightarrow \exists E \subset X\,(A \bowtie \neg E \wedge E \bowtie B)$.

This, too, is a condition that is satisfied by a large variety of apartness spaces, including metric spaces. Note that **EF** implies **B4**$_s$. If (X, τ) is a topological space satisfying

$$\forall x \in X \forall U \in \tau\,(x \in U \Rightarrow \forall y \in X\,(x \neq y \vee y \in U)), \tag{7}$$

then the pre-apartness defined at (3) satisfies the Efremovič condition.

3 Apartness and Topology

A subset S of a point–set pre-apartness space $(X, \bowtie)$ is said to be **nearly open** if it can be written as a union of apartness complements: that is, if there exists a family $(A_i)_{i \in I}$ such that $S = \bigcup_{i \in I} -A_i$. The nearly open sets form a topology—the **apartness topology**—on X for which the apartness complements form a basis.

Now consider a topological space (X, τ). We define the **topological point–set pre-apartness** corresponding to τ by

$$x \bowtie_\tau A \Longleftrightarrow \exists U \in \tau\ (x \in U \subset {\sim}A). \tag{8}$$

Taken with this pre-apartness, X becomes a **topological pre-apartness space**. The pre-apartness satisfies **A5** if and only if τ satisfies condition (7). In that case we call $(X, \bowtie_\tau)$ a **topological apartness space.**

If X is a pre-apartness space satisfying **B4**$_s$ and **A5**, then for all subsets A and B of X,

$$A \bowtie B \Longleftrightarrow \overline{A} \bowtie \overline{B},$$

where the overhead bar denotes closure with respect to the apartness topology on X.

In a topological pre-apartness space every nearly open set is open. It follows that if $(X, \bowtie)$ is a pre-apartness space with apartness topology $\tau_{\bowtie}$, then the topological pre-apartness corresponding to $\tau_{\bowtie}$ is just the original pre-apartness $\bowtie$ on X.

We say that a topological pre-apartness space is **topologically consistent** if every open subset of X is nearly open. Classically, every open set A in a topological pre-apartness space satisfies $A = -{\sim}A$, from which it follows that the space is topologically consistent. Constructively, although $A \subset -{\sim}A$ holds for an open set A, an example due to Jeremy Clarke (see [7]) shows that if every topological apartness space is topologically consistent, then the law of excluded middle holds. However, if the topological pre-apartness space (X, τ) is locally decomposable, then it is topologically consistent. In particular, every metric space is topologically consistent.

4 Point–Set Pre-Apartness and Continuity

As we are about to see, intuitionistic logic distinguishes between various types of continuity that are classically equivalent.

Let $f : X \to Y$ be a mapping between pre-apartness spaces. We say that f is

▷ **continuous** if

$$\forall x \in X \forall A \subset X \, (f(x) \bowtie f(A) \Rightarrow x \bowtie A) \, ;$$

▷ **topologically continuous** if $f^{-1}(S)$ is nearly open in X for each nearly open $S \subset Y$.

It is almost trivial that the composition of continuous functions is continuous, and that the restriction of a continuous function to an apartness subspace of its domain is continuous. Analogous remarks hold for topologically continuous ones.

Proposition 6 *A topologically continuous mapping between apartness spaces is continuous.*

Proof Let f be a topologically continuous map of an apartness space X into an apartness space Y. Consider $x \in X$ and $A \subset X$ such that $f(x) \bowtie f(A)$, and write

$$\Omega = f^{-1} \left(-f(A) \right) .$$

Since $-f(A)$ is nearly open, $\Omega = \bigcup_{i \in I} -A_i$ for some family of sets A_i. Choose $i \in I$ such that $x \in -A_i$. Note that $A \subset \neg\Omega$: for if $z \in A \cap \Omega$, then $f(z) \in f(A) \cap -f(A)$, which is absurd. Now, it can be shown that since Ω is nearly open, $\neg\Omega = \sim\Omega$. Hence

$$A \subset \neg\Omega = \sim\Omega \subset \sim - A_i$$

and therefore $-A_i \subset \sim A$. Applying **A4**, we now see that $x \bowtie A$ ■.

The converse, which holds classically, does not hold constructively. To see this, we need only consider Clarke's example of a topological space (X, τ) that is not provably topologically consistent. If f is the identity mapping from $(X, \tau_{\bowtie})$ to (X, τ), where $\tau_{\bowtie}$ denotes the apartness topology associated with the topological apartness $\bowtie$ on X, then f is continuous (in the apartness sense); however, if f is topologically continuous, then every open set in X is nearly open, which implies the law of excluded middle.

In order to obtain a partial converse to Proposition 6, we introduce the following **weak nested neighbourhoods property** for an apartness space X:

WNN $\quad x \in -T \Rightarrow \exists R \subset X (x \in -R \wedge (\neg R \subset -T)).$

Proposition 7 *Let X be a pre-apartness space, and Y a pre-apartness space satisfying WNN. Then every continuous function $f : X \to Y$ is topologically continuous.*

Proof Let $S \subset Y$ and $x_0 \in f^{-1}(-S)$—that is, $f(x_0) \in -S$. It will suffice to construct $T \subset X$ such that $x_0 \in -T$ and $f(-T) \subset -S$. To do this, we use the property **WNN** of Y to produce a subset R of Y such that $f(x_0) \in -R$ and $\neg R \subset -S$. Defining $T = f^{-1}(R)$, we see from the continuity of f that $x_0 \bowtie T$ and therefore $x_0 \in -T$. Moreover, for each $x \in -T$ we have $f(x) \notin R$; whence $f(x) \in -S$. ■

Note that condition **WNN** is a simple consequence of local decomposability.

5 Uniform Spaces

We now introduce a fundamental example of an apartness space. Let X be a inhabited set with an inequality $\neq$, and let U, V be subsets of the Cartesian product $X \times X$. We define certain associated subsets as follows:

$$\begin{aligned} U \circ V &= \{(x, y) : \exists z \in X((x, z) \in U \wedge (z, y) \in V)\}, \\ U^1 &= U, \quad U^{n+1} = U \circ U^n \quad (n = 1, 2, \ldots), \\ U^{-1} &= \{(x, y) : (y, x) \in U\}, \\ U[x] &= \{y \in X : (x, y) \in U\}. \end{aligned}$$

We say that U is **symmetric** if $U = U^{-1}$.

Let $\mathcal{U}$ be a family of inhabited subsets of $X \times X$, and for $S, T \subset X$ define

$$S \bowtie T \Longleftrightarrow \exists U \in \mathcal{U}\,(S \times T \subset {\sim}U).$$

If $\mathcal{U}$ has certain properties which we now present, then the set–set relation $\bowtie$ becomes a pre-apartness on X.

We say that $\mathcal{U}$ is a **uniform structure**, or a **uniformity**, on X if the following conditions hold:

U1

(i) Every finite intersection of sets in $\mathcal{U}$ belongs to $\mathcal{U}$.
(ii) Every subset of $X \times X$ that contains a member of $\mathcal{U}$ belongs to $\mathcal{U}$.

U2 For all $x, y \in X$, $x \neq y$ if and only if there exists $U \in \mathcal{U}$ such that $(x, y) \notin U$.
U3 For each $U \in \mathcal{U}$, $U^{-1} \in \mathcal{U}$ and there exists $V \in \mathcal{U}$ such that $V^2 \subset U$.
U4 For each $U \in \mathcal{U}$ there exists[5] $V \in \mathcal{U}$ such that $X \times X = U \cup \neg V$.

[5] Classically, property U4 always holds with $V = U$. It appears to be important to postulate it in the constructive theory.

The elements of $\mathcal{U}$ are called the **entourages** of (the uniform structure on) X, and the pair $(X, \mathcal{U})$—or simply X itself—is called a **uniform space**.

A metric space (X, ρ) is a uniform space in which the uniformity consists of all subsets of $X \times X$ that contain sets of the form $\{(x, y) : \rho(x, y) \leqslant \varepsilon\}$ for some $\varepsilon > 0$.

If U is an entourage of a uniform space $(X, \mathcal{U})$, then for all $x, y \in X$, either $x \neq y$ or $(x, y) \in U$. It follows that U contains the **diagonal**

$$\Delta = \{(x, x) : x \in X\}.$$

of X.

For each positive integer n we define an n**-chain of entourages** of X to be an n-tuple $(U_1, \ldots, U_n)$ of entourages such that $U_k^2 \subset U_{k-1}$ and $X \times X = U_{k-1} \cup {\sim}U_k$ for each $k \geqslant 2$. It can be shown that for each $U \in \mathcal{U}$ and each positive integer n there exists an n-chain $(U_1, \ldots, U_n)$ of entourages with $U_1 = U$ and each of $U_2, \ldots, U_n$ symmetric.

A given uniform structure $\mathcal{U}$ on X gives rise to the corresponding **uniform topology** $\tau_{\mathcal{U}}$, in which the sets $U[x]$, with $U \in \mathcal{U}$, are the neighbourhoods of the point x. This topology gives rise to a point–set apartness defined by

$$\textbf{apart}\,(x, S) \Longleftrightarrow \exists U \in \tau_{\mathcal{U}}\,(x \in U \subset {\sim}S)\,. \tag{9}$$

On the other hand, the uniform space X carries a **uniform apartness** defined by

$$S \bowtie T \Longleftrightarrow \exists U \in \mathcal{U}\,(S \times T \subset {\sim}U)\,. \tag{10}$$

It turns out that the corresponding point–set apartness is just the one, given at (9), corresponding to the uniform topology on X. The uniform apartness satisfies the Efremovič condition and hence $\mathbf{B4}_s$.

Given a general apartness space $(X, \bowtie)$, we say that a uniformity $\mathcal{U}$ on X is **compatible** with, or **induces**, the given apartness if the apartness corresponding to $\mathcal{U}$ is the same as $\bowtie$ —in other words if (10) holds. In the classical theory of proximity spaces, every proximity space satisfying the Efremovič condition has uniform structures that are compatible with the denial apartness (the one in which two sets are apart if and only if they are not near each other); see [17], p. 71. The following strange result helps us to show that things are not so simple constructively.

Proposition 8 *If there exists a pre-apartness space X such that*

(i) $A \bowtie B \Rightarrow \forall x \in X\,(x \notin A \vee x \notin B)$ *and*
(ii) *any two disjoint subsets of a singleton are apart (that is, if $x \in X, A \subset \{x\}, B \subset \{x\}$, and $A \cap B = \varnothing$, then $A \bowtie B$),*

then the ***weak law of excluded middle***

$$\neg P \vee \neg\neg P$$

holds.

Proof Suppose there exists such an apartness space X. Fixing $\xi \in X$, consider any syntactically correct statement P, and define

$$A = \{x \in X : x = \xi \wedge P\},$$
$$B = \{x \in X : x = \xi \wedge \neg P\}.$$

Then A and B are disjoint subsets of $\{\xi\}$, so, by hypothesis (ii), $A \bowtie B$. By hypothesis (i), either $\xi \notin A$, in which case $\neg P$ holds, or else $\xi \notin B$ and therefore $\neg\neg P$ holds. ■

The somewhat eccentric hypothesis (ii) in Proposition 8 holds classically for any apartness space: for if A, B are disjoint subsets of a singleton, then classically either A or B is empty, so axiom **B1** applies. The hypothesis holds constructively if the apartness on X is induced by a uniform structure $\mathcal{U}$: for if $\xi \in X$, and A, B are disjoint subsets of $\{\xi\}$, then, taking any $U \in \mathcal{U}$, we have $A \times B = \varnothing \subset \sim U$; whence $A \bowtie B$.

If X is a metric space, $x \in X$, and $S \subset X$, then we write $\rho(x, S) > 0$ to signify that there exists $r > 0$ such that $\rho(x, s) \geqslant r$ for all $s \in S$."

Corollary 9 *Let ρ denote the standard metric on $\mathbb{R}$, and let $\bowtie$ be the symmetric set–set pre-apartness defined on $\mathbb{R}$ by*

$$A \bowtie B \Longleftrightarrow \forall x \in \mathbb{R}\,(\rho(x, A) > 0 \vee \rho(x, B) > 0).$$

If $\bowtie$ is induced by a uniform structure, then the weak law of excluded middle holds.

Proof It is clear that $\bowtie$ satisfies condition (i) of the Proposition 8. If it is induced by a uniform structure, then, as the remark preceding this corollary explains, so is hypothesis (ii) of Proposition 8. The result now follows from that proposition. ■

Classically, the pre-apartness defined in the Corollary 9 satisfies the Efremovič condition and so has compatible uniformities. The corollary shows that, in a sense, the constructive theory of apartness spaces is strictly bigger than that of uniform spaces. Moreover, the compact metric subspace $[0, 1]$ of $\mathbb{R}$ has an apartness—the one induced by the set–set apartness in Corollary 9—which induces the original metric topology. However, without the weak law of excluded middle we cannot prove that this is the same as the metric apartness. This contrasts with the classical situation in which any compact topological space has a unique proximity structure compatible with the original compact topology.

Corollary 10 *Let X be an inhabited set with the denial inequality. Then*

$$S \bowtie T \Longleftrightarrow S = \varnothing \vee T = \varnothing \tag{11}$$

defines a set–set apartness on X that satisfies the Efremovič condition. If this apartness is induced by a uniform structure, then the weak law of excluded middle holds.

Proof It is routine to verify that the point–set relation $\bowtie_0$ defined by

$$x \bowtie_0 T \Longleftrightarrow T = \varnothing$$

satisfies **A1**–**A5**. Moreover, it is locally decomposable: for if $x \bowtie_0 U$ and therefore $U = \varnothing$, then $X = {\sim}U \cup U$. Noting that

$$\forall x \in X\, (x \bowtie S \vee x \bowtie T) \Longleftrightarrow (S = \varnothing \vee T = \varnothing)\,,$$

we easily show that the set–set relation defined at (11) is a symmetric set–set apartness. As in the preceding proof, we also see that if it is induced by a uniform structure, then the weak law of excluded middle holds. Finally, consider S, T with $S \bowtie T$. If $S = \varnothing$, then $S \bowtie_0 \neg\varnothing$ and $\varnothing \bowtie_0 T$. If $T = \varnothing$, then $S \bowtie_0 \neg X$ and $X \bowtie_0 T$. Thus $\bowtie_0$ satisfies **EF**. ■

This corollary shows even more clearly than does Corollary 9 that our theory of apartness spaces is larger than that of uniform spaces. For it provides an example of a set–set apartness for which, without the weak law of excluded middle, we cannot produce an inducing uniform structure *even though the apartness satisfies the Efremovič condition.*

We emphasise that the set–set pre-apartnesses in the preceding two corollaries should not be regarded as pathological: classically each can be described by saying that $A \bowtie B$ if and only if the closures of A and B in the metric topology on $\mathbb{R}$ are disjoint. The negation of this pre-apartness is a standard classical example of a proximity.

6 Strong and Uniform Continuity

We now introduce an important and natural continuity property of a mapping $f : X \to Y$ between set–set apartness spaces. We say that f is **strongly continuous** if

$$\forall S, T \subset X\, (f(S) \bowtie f(T) \Rightarrow S \bowtie T)\,.$$

Strongly continuous mappings are the morphisms in the category of set–set apartness spaces. Compositions of such mappings are also strongly continuous. Clearly, strong continuity implies continuity.

A mapping f between uniform (apartness) spaces $(X, \mathcal{U})$ and $(Y, \mathcal{V})$ is said to be **uniformly continuous** if

$$\forall V \in \mathcal{V} \exists U \in \mathcal{U} \forall x, x' \in X\, \left(\left(x, x'\right) \in U \Rightarrow \left(f(x), f(x')\right) \in V\right).$$

Uniform continuity implies strong continuity. The converse holds if $f(X)$ is totally bounded—that is, for each $V \in \mathcal{V}$ there exist finitely many points $x_1, \ldots, x_n$ of X such that

$$f(X) = \bigcup_{i=1}^{n} \{f(x) : x \in X,\ \ (f(x), f(x_i)) \in V\}.$$

In order to discuss the relation between strong and uniform continuity more generally, we introduce a new notion. We say that two sequences $(x_n)_{n\geqslant 1}$, $\left(x_n'\right)_{n\geqslant 1}$ in a uniform space $(X, \mathcal{U})$ are **eventually close** if

$$\forall U \in \mathcal{U} \exists N \forall n \geqslant N \ \left(\left(x_n, x_n'\right) \in U\right).$$

A mapping f of X into a uniform space Y is **uniformly sequentially continuous** if the sequences $(f(x_n))_{n\geqslant 1}$, $\left(f(x_n')\right)_{n\geqslant 1}$ are eventually close in Y whenever $(x_n)_{n\geqslant 1}$, $\left(x_n'\right)_{n\geqslant 1}$ are eventually close in X. The fundamental result about strong continuity is the following, which we state without its complicated proof.

Theorem 11 *A strongly continuous mapping $f : X \to Y$ between uniform spaces is uniformly sequentially continuous.*

In proving this theorem we were led to a new proof-technique which turned out to have several applications in the theory of apartness; see [4]. Moreover, we showed that a key classical result, used by us to rule out a certain case en route to proving Theorem 11, has a much more natural proof than the standard one, which uses the full power of the law of excluded middle.

It is shown in [8] that, in the case where X, Y are metric spaces, Theorem 11 cannot be extended to produce the uniform continuity of f unless we employ a principle that, while valid in the intuitionistic, recursive, and classical models of constructive mathematics, appears not to be derivable with only intuitionistic logic and dependent choice. However, under stronger hypotheses we can obtain the conclusion that a strongly continuous function between uniform spaces is not just uniformly sequentially continuous but uniformly continuous. In particular, every strongly continuous mapping from a uniform space onto a totally bounded uniform space is uniformly continuous.

7 Apartness on Function Spaces

In this section we consider pre-apartness structures on Y^X, where X is an inhabited set and Y is a uniform space. The inequality on Y^X is defined in the normal way by

$$f \neq g \Longleftrightarrow \exists x \in X\,(f(x) \neq g(x)).$$

For all $A \subset X$ and $B \subset Y$ we define

$$U_{A,B} = \left\{f \in Y^X : f(A) \bowtie B\right\}.$$

The sets $U_{A,B}$, with $A \subset X$ and $B \subset Y$, form a subbase of a topology τ_p on Y^X called the **topology of proximal convergence**. We obtain from this topology a point–set pre-apartness on Y^X by setting $f \bowtie_{Y^X} S$ if and only if there exist finitely many subsets $A_1, \ldots, A_m$ of X and finitely many subsets $B_1, \ldots, B_m$ of Y such that

$$f \in \bigcap_{i=1}^{m} U_{A_i,B_i} \subset \sim S.$$

It is easy to show that $\bowtie_{Y^X}$ is a pre-apartness on Y^X. Moreover, if $f, g \in Y^X$ and $f \bowtie_{Y^X} \{g\}$, then $f \in \sim\{g\}$, so $f \neq g$. However, a simple Brouwerian example shows that we cannot prove constructively that **A5** holds for $\bowtie_{Y^X}$ even when Y is a two–point apartness space.

A net $(f_n)_{n\in D}$ of elements of the function space Y^X is said to be **proximally convergent** to an element f of Y^X if[6]

$$\forall A \subset X \forall B \subset Y \left(f(A) \bowtie B \Rightarrow \exists N \in D \forall n \geqslant N \left(f_n(A) \bowtie B\right)\right).$$

It turns out that $(f_n)_{n\in D}$ converges to $f \in Y^X$ proximally if and only if it converges to f in the topology τ_p. On the other hand, proximal convergence in Y^X implies $\bowtie_{Y^X}$-convergence; but the converse holds if and only if $\left(Y^X, \tau_p\right)$ is topologically consistent.

When Y carries a uniform structure $\mathcal{U}$, we have the problem of linking proximal convergence in Y^X with uniform convergence.

For each $U \in \mathcal{U}$ we define

$$W_X(U) = \left\{(f, g) \in Y^X \times Y^X : \forall x \in X \left((f(x), g(x)) \in U\right)\right\}.$$

The set

$$\mathcal{W} = \{W_X(U) : U \in \mathcal{U}\}$$

then satisfies the standard classical axioms for a uniform structure on Y^X. However, if $\left(Y^X, \mathcal{W}\right)$ satisfies **U4**, then LPO holds; so for us $\mathcal{W}$ is just a **pre-uniform structure**: it satisfies **U1–U3**. Uniform convergence in Y^X is equivalent to $\bowtie_{\mathcal{W}}$-convergence.

The pre-uniform structure $\mathcal{W}$ gives rise to a topology $\tau_{\mathcal{W}}$ on Y^X, in which the sets

$$W_X(U)[f] = \left\{g \in Y^X : (f, g) \in W_X(U)\right\}$$

form a base of neighbourhoods of f. A net $(f_n)_{n\in D}$ converges to f in this topology if and only if for each $U \in \mathcal{U}$ there exists $N \in D$ such that $(f, f_n) \in W_X(U)$—that is,

[6] For more on proximal convergence see [22, 23].

$$\forall x \in X\,((f(x), f_n(x)) \in U)$$

—for all $n \succcurlyeq N$. In other words, convergence with respect to $\tau_{\mathcal{W}}$ is just what we already know as uniform convergence.

Corresponding to the pre-uniform structure $\mathcal{W}$ is a relation $\bowtie_{\mathcal{W}}$ on subsets of Y^X, defined by

$$S \bowtie_{\mathcal{W}} T \Longleftrightarrow \exists U \in \mathcal{U}\,(S \times T \subset \neg W_X(U)).$$

This is easily shown to be a **weak pre-apartness**: that is, it satisfies **B1**, **B3**, **B4**, and

B2$_w$ $\quad A \bowtie B \Rightarrow A \cap B = \varnothing$.

A Brouwerian example shows that for the relation $\bowtie_{\mathcal{W}}$ we cannot prove even this weak version of **A5**,

$$x \bowtie S \Rightarrow \forall y \in X\,(\neg(x = y) \vee y \bowtie S), \tag{12}$$

let alone **B5**.

What, if any, are the connections between τ_p and $\tau_{\mathcal{W}}$, between $\bowtie_{Y^X}$ and $\bowtie_{\mathcal{W}}$, and between various notions of convergence associated with those structures on Y^X? The topology $\tau_{\mathcal{W}}$ of uniform convergence on Y^X is finer than the topology of proximal convergence; so uniform convergence of a net in Y^X implies proximal convergence. If Y is a totally bounded uniform space, then the topologies τ_p and $\tau_{\mathcal{W}}$ coincide, so proximal convergence of nets in Y^X is then equivalent to uniform convergence.

We always have $\bowtie_{Y^X} \subset \bowtie_{\mathcal{W}}$. If the reverse inclusion holds, then $\bowtie_{Y^X}$-convergence, uniform convergence, and proximal convergence are equivalent. Since there is a classical example, due to Nachman [16], showing that proximal convergence need not imply uniform convergence, we cannot guarantee that $\bowtie_{\mathcal{W}} \subset \bowtie_{Y^X}$ in general. Note that if $\bowtie_{\mathcal{W}} \subset \bowtie_{Y^X}$, then the space $\left(Y^X, \tau_p\right)$ is topologically consistent. For more on these matters, see [5].

8 Point-Free Apartness Spaces

In keeping with the modern trend towards "point-free" theories in topology and analysis, we have recently started to examine apartness on lattices, of which we now present a rough summary.[7]

Let $\mathfrak{L}$ be a lattice with partial order $\leqslant$ and the usual total binary operations $\vee$ ("join") and $\wedge$ ("meet"). We assume that $\mathfrak{L}$ also has a top element 1 and a bottom element 0 : thus $0 \leqslant x \leqslant$ for each $x \in \mathfrak{L}$. We further assume that $\mathfrak{L}$ is **complemented**

[7] In connection with this, note the recent work of Curi [11] and of Palmgren & Schuster [18].

that is, there is a unary function $\sim : \mathfrak{L} \to \mathfrak{L}$ of **orthocomplementation**, satisfying the following properties.

- $x \wedge {\sim}x = 0$
- $x \wedge {\sim}{\sim}x = x$
- ${\sim}(x \vee y) = {\sim}x \wedge {\sim}y$
- $x \leqslant y \Rightarrow {\sim}y \leqslant {\sim}x$

We say that $\mathfrak{L}$ is

- **distributive** if

$$\forall x, y, z \in \mathfrak{L}\,(x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z))\,;$$

- **modular** if

$$\forall x, y, z\,(x \leqslant z \Rightarrow x \wedge (y \vee z) = (x \wedge y) \vee z)\,.$$

Every distributive lattice is modular. In a modular complemented lattice $\mathfrak{L}$,

$$\forall x \forall y\,(x \leqslant {\sim}x \vee y \Rightarrow x \leqslant y)\,,$$

$$\forall x\,(x = {\sim}{\sim}x) \Longleftrightarrow \forall x \in \mathfrak{L}\,(x \vee {\sim}x = 1)\,,$$

and

$$\forall x \forall y\,({\sim}x \vee {\sim}y \leqslant {\sim}(x \wedge y))\,.$$

In connection with the second of these displays, note that, in view of the situation in the set model (where $\mathfrak{L}$ is the set of subsets of an inhabited set X, with the usual lattice structure), we cannot hope to prove, for a general $\mathfrak{L}$, either that

$$\forall x\,(x = {\sim}{\sim}x)$$

or that

$$\forall x \in \mathfrak{L}\,(x \vee {\sim}x = 1)\,.$$

We define the **join and meet of an arbitrary family** $(x_i)_{i \in I}$ of elements of a lattice in the usual way:

$$x = \bigvee_{i \in I} x_i \Longleftrightarrow \forall i \in I\,(x_i \leqslant x)\ \&\ \forall y\,(\forall i \in I\,(x_i \leqslant y) \Rightarrow x \leqslant y)\,,$$
$$x = \bigwedge_{i \in I} x_i \Longleftrightarrow \forall i \in I\,(x \leqslant x_i)\ \&\ \forall y\,(y \leqslant x \Rightarrow \forall i \in I\,(y \leqslant x_i))\,.$$

These elements need not exist when I is an infinite index set.

Classically, every set comes equipped with an inequality relation—namely, the denial inequality. In the constructive setting we postulate that our lattice $\mathfrak{L}$ has a unary **habitation relation** hab, whose axiomatic requirements mirror those of inhabitedness for sets:

H1 $\mathrm{hab}(x) \Rightarrow \neg(x = 0)$
H2 $(\mathrm{hab}(x) \wedge (x \leqslant y)) \Rightarrow \mathrm{hab}(y)$
H3 The **join-existential property:** For any family $(x_i)_{i \in I}$ of elements of $\mathfrak{L}$, if $\bigvee_{i \in I} x_i$ and exists and $\mathrm{hab}\left(\bigvee_{i \in I} x_i\right)$, then there exists $i \in I$ such that $\mathrm{hab}(x_i)$.
H4 $\mathrm{hab}(1)$.

If $\mathrm{hab}(x)$, we say that x is an **inhabited** element of $\mathfrak{L}$.

If $\mathfrak{L}$ is complemented and has a habitation relation, we define a binary relation $\neq$ on $\mathfrak{L}$ by

$$x \neq y \Longleftrightarrow (\mathrm{hab}(x \wedge \sim y) \vee \mathrm{hab}(\sim x \wedge y)).$$

Then

$$x \neq 0 \Longleftrightarrow \mathrm{hab}(x).$$

Also,

$$x \neq y \Longleftrightarrow y \neq x,$$
$$x \neq y \Rightarrow \neg(x = y),$$

so the relation $\neq$ is a genuine inequality relation. Note that since $1 \wedge \sim 0 = 1$, it follows from **H4** that $0 \neq 1$.

We define the **logical complement**[8] of an element x of a lattice $\mathfrak{L}$ to be

$$\neg x = \bigvee \{y \in \mathfrak{L} : y \wedge x = 0\},$$

if that join exists. We say that $\mathfrak{L}$ **has logical complements** or is **with logical complements** if each element of $\mathfrak{L}$ has a logical complement. The logical complement of x is the unique element z of $\mathfrak{L}$ such that

$$z \wedge x = 0 \;\&\; \forall y \in \mathfrak{L}\, (y \wedge x = 0 \Rightarrow y \leqslant z).$$

We note two lemmas about logical complements:

▷ If $x \leqslant y$, then $\neg y \leqslant \neg x$.
▷ In a modular lattice, $\neg(x \vee y) = \neg x \wedge \neg y$.

[8] In classical lattice theory, the logical complement is called the *pseudo-complement.*

8.1 A-Frames

A lattice $\mathfrak{L}$ is said to be **complete** if the join, and hence the meet, exists for any family of elements of $\mathfrak{L}$. Each element of a complete lattice has a logical complement.

A complemented lattice $\mathfrak{L}$ is called a **frame** if

- it is complete,
- it has a habitation relation, and
- for all $x \in \mathfrak{L}$ and all families $(u_i)_{i \in I}$ of elements of $\mathfrak{L}$,

$$x \wedge \bigvee_{i \in I} u_i = \bigvee_{i \in I} (x \wedge u_i)$$

—in other words, $\wedge$ is **infinitely distributive** over $\vee$.

We now introduce two fundamental relations on a frame $\mathfrak{L}$: a symmetric binary relation $\bowtie$ and an associated unary relation $-$, where for each $x \in \mathfrak{L}$,

$$-x = \bigvee \{y \in \mathfrak{L} : y \bowtie x\}.$$

For $\bowtie$ to be a **frame pre-apartness** we require that the following axioms be satisfied:

AL1 $1 \bowtie 0$
AL2 $x \bowtie y \Rightarrow x \leqslant \sim y$
AL3 $x \bowtie (y \vee z) \Longleftrightarrow (x \bowtie y \;\&\; x \bowtie z)$

Note that we do not require symmetry of the relation $\bowtie$ in the context of a lattice. If $x \bowtie y$, we say that x and y are **apart**, and we call $-x$ the **apartness complement** of x. Taken with a frame apartness, $\mathfrak{L}$ becomes an **apartness frame**, or an **a-frame.** In an a-frame we have $-0 = 1$, $-1 = 0$, $-x \leqslant \sim x$, and $-(x \vee y) = -x \wedge -y$; also, if $x \leqslant y$ and $y \bowtie z$, then $x \bowtie z$.

The following **Lodato property** is the translation of the set–set apartness axiom **B4** into the context of a-frames:

AL4 $-x \leqslant \sim y \Rightarrow -x \leqslant -y$.

If this holds, we say that $\bowtie$ is a **Lodato pre-apartness** on $\mathfrak{L}$, and that $\mathfrak{L}$ is a **Lodato a-frame**.

The obvious example of an a-frame is the set of subsets of a set–set apartness space, with hab (S) meaning "there exists an element of S". Another example occurs when we have a topological space (X, τ) with an inequality relation $\neq$, and we consider the set $\mathfrak{L}$ of all **regular open sets:** that is, subsets U of X such that $U = \overline{U}^{\circ}$. In this example, the complement of U is defined to be $(\sim U)^{\circ}$, and the lattice ordering is by set inclusion. The lattice operations are then given by $U \vee V = \overline{U \cup V}^{\circ}$ and $U \wedge V = U \cap V$. The apartness that turns $\mathfrak{L}$ into an a-frame is defined by

$$U \bowtie V \Leftrightarrow U \subset (\sim V)^{\circ}.$$

This lattice may have applications in a constructive version of the “region connection calculus”, as studied by certain theoretical computer scientists [21].

Everything we have done so far with lattices is point-free. However, we can introduce points, or “atoms”, into our theory as follows: an **atom** of a lattice $\mathfrak{L}$ to be an element x with the properties

$$x \neq 0 \ \& \ \forall y \, (0 \neq y \leqslant x \Rightarrow y = x).$$

Thus atoms correspond to singleton subsets in the set-set apartness model.

If $\mathfrak{L}$ is a frame, x is an atom, and $x \leqslant \bigvee_{i \in I} u_i$, then there exists i such that $x \leqslant u_i$. If $\mathfrak{L}$ is an a-frame and x is an atom such that $x \leqslant -y$, then $x \bowtie y$.

8.2 Topology-Like Structures

We now look at some lattice analogues of topological notions. Consider any frame $\mathfrak{L}$ with a family τ of elements satisfying the following three properties:

TL1 $0 \in \tau$ and $1 \in \tau$.
TL2 If $(u_i)_{i \in I}$ is a family of elements of τ, then $\bigvee_{i \in I} u_i$ belongs to τ.
TL3 If $u, v \in \tau$, then $u \wedge v \in \tau$.

We call τ a **topology-like structure**, or a **t-structure**, on $\mathfrak{L}$, the pair $(\mathfrak{L}, \tau)$ a **topological frame**, and the elements of τ the corresponding **open elements** of $\mathfrak{L}$.

If $\mathfrak{L}$ is an a-frame, then, by analogy with the set-set case, it makes sense to define the **nearly open elements** to be those of the form $\bigvee_{i \in I} - u_i$. The set $\tau_{\mathfrak{L}}$ of nearly open sets is a t-structure on $\mathfrak{L}$.

Given a topological frame $(\mathfrak{L}, \tau)$, we define a relation $\bowtie_\tau$ on $\mathfrak{L}$ as follows:

$$x \bowtie_\tau y \Longleftrightarrow \exists u \in \tau \, (x \leqslant u \leqslant \sim y).$$

We also define

$$-_\tau x = \bigvee \{z \in \mathfrak{L} : z \bowtie_\tau x\}.$$

Then $\bowtie_\tau$ is a Lodato pre-apartness, the **topological pre-apartness**, on $\mathfrak{L}$.

For any element x of a topological frame, the **interior** of x is defined by

$$x^{\circ} = \bigvee \{u \in \tau : u \leqslant x\}.$$

Then $x^\circ \in \tau$, by **TL2**; and the definition of "join" shows that $x^\circ \leqslant x$. If $x \in \tau$, then since $x \leqslant x$, we also have $x \leqslant x^\circ$ and therefore $x = x^\circ$.

Now that we have a lattice pre-apartness $\bowtie_\tau$ and an apartness complement $-_\tau$ on a topological frame $(\mathfrak{L}, \tau)$, we have the corresponding nearly open elements of $\mathfrak{L}$: namely, the joins of elements of the form $-_\tau x$. Every nearly open element is open; but, as the set model shows, we cannot expect to prove that every open element of $\mathfrak{L}$ is nearly open. If that property does hold, then we call $\mathfrak{L}$ **topologically consistent.**

We say that an a-frame $\mathfrak{L}$ is **locally decomposable** if

$$\forall x \in \mathfrak{L}\left(-x = \bigvee \{-y \in \mathfrak{L} : -x \vee y = 1\}\right).$$

In the set model this property is equivalent to that in axiom **B5**, and so turns the pre-apartness into an apartness. As in the set model, local decomposability is enough to guarantee topological consistency. Note also that local decomposability entails

$$\forall x \in \mathfrak{L}\left(-x = \bigvee \{y \in \mathfrak{L} : -x \vee -y = 1\}\right), \tag{13}$$

which is a lattice version of the point–set property **A5.**

8.3 Frame Maps and Continuity

A mapping $f : \mathfrak{L} \to \mathfrak{M}$ between a-frames is called a **frame map** if

- ▷ for each family $(x_i)_{i\in I}$ of elements of $\mathfrak{L}$, $f\left(\bigvee_{i\in I} x_i\right) = \bigvee_{i\in I} f(x_i)$, and
- ▷ for all $x \in \mathfrak{L}$, $x \neq 0$ if and only if $f(x) \neq 0$.

Such a map is order-preserving: if $a \leqslant b$, then $a \vee b = b$, so $f(a) \vee f(b) = f(a \vee b) = f(b)$ and therefore $f(a) \leqslant f(b)$. It follows that

$$f\left(\bigwedge_{i\in I} x_i\right) \leqslant \bigwedge_{i\in I} f(x_i)$$

for all families $(x_i)_{i\in I}$ of elements of $\mathfrak{L}$.

Let $f : \mathfrak{L} \to \mathfrak{M}$ be a frame map between a-frames. For each $v \in \mathfrak{M}$ define

$$f^{-\infty}(v) = \bigvee \{x \in \mathfrak{L} : f(x) \leqslant v\}.$$

We say that f is

- **continuous** if

$$\forall s, t \in \mathfrak{M} \left(s \leqslant -t \Rightarrow f^{-\infty}(s) \leqslant -f^{-\infty}(t) \right);$$

- **topologically continuous** if

$$\forall v \in \tau_{\mathfrak{M}} \left(f^{-\infty}(v) \in \tau_{\mathfrak{L}} \right),$$

where, for example, $\tau_{\mathfrak{L}}$ donates the t-structure on $\mathfrak{L}$;

- **strongly continuous** if

$$\forall s, t \in \mathfrak{M} \left(s \bowtie t \Rightarrow f^{-\infty}(s) \bowtie f^{-\infty}(t) \right).$$

The element $f^{-\infty}(x)$ plays the role of the inverse image of a set in the set model. For example, in that model topological continuity of a mapping f between set–set apartness spaces X and Y means that $f^{-1}(V)$ is nearly open in X for each nearly open $V \subset Y$. But

$$\begin{aligned} f^{-1}(V) &= \{x \in X : f(x) \in V\} \\ &= \bigcup \{U \subset X : f(U) \subset V\}, \end{aligned}$$

which is precisely the counterpart of $f^{-\infty}(v)$ in the definition of “topologically continuous” for a frame map. We have used the special notation $f^{-\infty}(v)$ in order to avoid confusion with $f^{-1}(v) = \{x \in \mathfrak{L} : f(x) = v\}$.

It is easy to see that the composition of two frame maps of the same continuity type is also of that continuity type.

Clearly, strong continuity implies continuity. What other relations do we have between our various notions of continuity? Here are two, without proofs.

Proposition 12 *Let $\mathfrak{L}$ be a Lodato a-frame that satisfies* (13) *and whose inequality is **zero-tight** in the sense that*

$$\forall x \in \mathfrak{L} \, (\neg (x \neq 0) \Rightarrow x = 0).$$

Then every topologically continuous frame map from $\mathfrak{L}$ to an a-frame is continuous.

We call a lattice $\mathfrak{L}$ **atomic** if

$$\forall_{x \in \mathfrak{L}} \left(x \neq 0 \Rightarrow x = \bigvee \{y \in \mathfrak{L} \ : \ y \text{ is an atom and } y \leqslant x\} \right).$$

Proposition 13 *Let $\mathfrak{L}$ and $\mathfrak{M}$ be a-frames, with $\mathfrak{L}$ atomic and $\mathfrak{M}$ locally decomposable. Let $f : \mathfrak{L} \to \mathfrak{M}$ be a continuous frame map which preserves atoms. Then f is topologically continuous.*

9 Concluding Remarks

Since its inception early in 2000 the theory of apartness between points and sets, and between sets and sets, has provided a firm foundation for constructive topology. In the foregoing we have merely outlined some of the fundamental ideas, without touching upon such matters as separation properties (other than **EF**), convergence and completeness, connectedness, or compactness. The last of these seems particularly hard to bring into our theory; since the only constructively viable notion of compactness in uniform spaces is that of total boundedness plus completeness, it is not at all clear how one might generalise total boundedness beyond uniform spaces to apartness spaces. The problem of finding a good definition of compactness for an apartness space is considered further in [9, 13].

The point-free theory of apartness in lattices is really in its infancy, or at least its pre-school stage. However, as we have indicated, even notions (such as local decomposability) which appear to depend on the existence of points can be lifted from the point–set context to that of an a-frame, thereby leading to results for frame mappings that are analogous to those for set–set mappings. Note that a theory of product apartness structures on frames is developed in [6].

Acknowledgments The authors thank the New Zealand Foundation for Science & Technology for supporting Luminiţa Vîţă as a Postdoctoral Research Fellow from 2002–2006.

References

1. P. Aczel and M. Rathjen, *Notes on Constructive Set Theory*, Report No. 40, Institut Mittag-Leffler, Royal Swedish Academy of Sciences, 2001.
2. E.A. Bishop, *Foundations of Constructive Analysis*, McGraw-Hill, New York, 1967.
3. D.S. Bridges and L.S. Vîţă, 'Apartness spaces as a framework for constructive topology', Ann. Pure Appl. Logic. **119** (1–3), 61–83, 2003.
4. D.S. Bridges and L.S. Vîţă, 'A general constructive proof technique', Elec. Notes in Theoretical Comp. Sci. **120**, 31–43, 2005.
5. D.S. Bridges and L.S. Vîţă, 'Pre-apartness structures on spaces of functions', J. Complexity **22**, 881–893, 2006.
6. D.S. Bridges, 'Product a-frames and proximity', Math. Logic Quarterly 54(1), 12–25, 2008.
7. D.S. Bridges and L.S. Vîţă, *Apartness Spaces*, research monograph, in preparation.
8. D.S. Bridges, H. Ishihara, P.M. Schuster, and L.S. Vîţă, 'Strong continuity implies uniform sequential continuity', Archive for Math. Logic **44**(7), 887–895, 2005.
9. D.S. Bridges, H. Ishihara, P.M. Schuster, and L.S. Vîţă, 'Apartness, compactness and nearness', Theoretical Comp. Sci. **405**, 3–10, 2008.
10. P. Cameron, J.G. Hocking, and S.A. Naimpally, 'Nearness—a better approach to continuity and limits', Amer. Math. Monthly, Sept. 1974, 739–745.
11. G. Curi, 'Notes on 'Apartness Spaces as Formal Spaces'', preprint, Università di Padova, Padova.
12. H. Diener, 'Generalising compactness', Math. Logic Quarterly **54**(1), 49–57, 2008.
13. R.J. Grayson, 'Concepts of general topology in constructive mathematics and in sheaves I', Ann. Math. Logic **20**, 1–41, 1981.

14. R.J. Grayson, ‘Concepts of general topology in constructive mathematics and in sheaves II’, Ann. Math. Logic **23**, 55–98, 1982.
15. R. Milošević, D.A. Romano, and M. Vinčić, ‘A basic separation on a set with apartness’, Publikacije Elektr. Fakult. Univ. u Beogradu **8**, 37-43, 1997.
16. L.J. Nachman, *Weak and Strong Constructions in Proximity Spaces,* Ph.D. dissertation, The Ohio State University, 1968.
17. S.A. Naimpally and B.D. Warrack, *Proximity Spaces*, Cambridge Tracts in Math. and Math. Phys. **59**, Cambridge at the University Press, 1970.
18. E. Palmgren and P.M. Schuster, ‘Apartness and formal topology’, New Zealand J. Math. **34**, 1–8, 2005.
19. G. Sambin, ‘Intuitionistic formal spaces—a first communication’, in: *Mathematical Logic and its Applications* (D. Skordev, ed.), 187–204, Plenum Press, New York, 1987.
20. G. Sambin, ‘Some points in formal topology’, Theoretical Computer Science **305**, 347–408, 2003.
21. J. Stell, ‘Boolean connection algebra: a new approach to the region connection calculus’, Artificial Intell. **122**, 111–136, 2002.
22. L.S. Vîţă, ‘On proximal convergence in uniform spaces’, Math Logic Quarterly, **49**(6), 550–552, 2003.
23. L.S. Vîţă, ‘Proximal and uniform convergence on apartness spaces’, Math. Logic Quarterly, **49**(3), 255–259, 2003.
24. F. Waaldijk, *Modern Intuitionistic Topology,* Ph.D. thesis, University of Nijmegen, Netherlands, 1996.

Relativization of Real Numbers to a Universe

Hajime Ishihara*

Abstract We discuss a relativization of real numbers to a universe given by a function algebra, and develop a tentative theory of relativized real numbers. We show that the class $\mathbf{R}(\mathcal{F}\text{PTIME})$ of real numbers, obtained by relativizing to the class $\mathcal{F}\text{PTIME}$ of polynomial time computable functions, is a proper subclass of the class $\mathbf{R}(\mathcal{E})$ of real numbers, obtained by relativizing to the class $\mathcal{E}$ of elementary functions. We show the Cauchy completeness of relativized real numbers, and that we can prove the (constructive or approximate) intermediate value theorem if our universe is closed under a closure condition used to characterize the polynomial time computable functions.

1 Introduction

A real number in the constructive theory [1, 2, 15] is defined as a Cauchy sequence of rational numbers, and, since we can code rational numbers by natural numbers, a real number is a sequence (or unary function) of natural numbers. Therefore the notion of real numbers is relative to a universe of functions of natural numbers.

In the constructive theory of real numbers, developed in [15, Chapter 5], we assume that a universe $\mathcal{U}$ satisfies a closure conditions expressing the fact that $\mathcal{U}$ is closed under "recursive in". As mentioned in [15, 5.4.6], the theory may be correct if we relativize the notion of sequence of natural numbers to some other universe of functions of natural numbers. For a smooth theory of real numbers, the universe has to contain some functions and to satisfy certain closure conditions.

H. Ishihara (✉)
School of Information Science, Japan Advanced Institute of Science and Technology,
Nomi, Ishikawa 923-1292, Japan
e-mail: Ishihara@jaist.ac.jp

*Partly supported by a Grant-in-Aid for Scientific Research (C) No.15500005 of Japan Society for the Promotion of Science.

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_9,

On the other hand, various classes of (total) functions on natural numbers have been investigated so far, such as the total recursive functions, the primitive recursive functions, the elementary functions, and the classes in the (extended) Grzegorczyk hierarchy etc. [14]. Computer scientists have focussed on classes, based on time or space constraint on a computation of a machine, such as the logarithmic space computable functions, the polynomial time computable functions, and the polynomial space computable functions etc. [9].

Some of such classes have been characterized by function algebras: a *function algebra* is the smallest class of functions containing a set of (initial) functions and closed under certain closure conditions. We refer the reader to [3] for a good exposition of function algebras, and [6, 12] for recent results.

In this paper, we discuss a relativization of real numbers to a universe given by a function algebra, and develop a tentative theory of relativized real numbers.

Although the relativization of real numbers to a universe is quite natural and seems quite promising, does it really make sense, especially for a small universe? For example, consider the class $\mathcal{F}\text{PTIME}$ of polynomial time computable functions and the class $\mathcal{E}$ of elementary functions. It is known that $\mathcal{F}\text{PTIME} \subsetneq \mathcal{E}$. Do the classes $\mathbf{R}(\mathcal{F}\text{PTIME})$ and $\mathbf{R}(\mathcal{E})$ of real numbers, obtained by relativizing to $\mathcal{F}\text{PTIME}$ and $\mathcal{E}$ respectively, give two distinct classes?

After reviewing the classes $\mathcal{F}\text{PTIME}$ and $\mathcal{E}$ in Section 2, we show, in Section 3, that $\mathbf{R}(\mathcal{F}\text{PTIME})$ is a *proper* subclass of $\mathbf{R}(\mathcal{E})$.

In Section 4, we develop a basic theory of relativized real numbers, and show the Cauchy completeness of relativized real numbers. Then, in the last section, we show that the (constructive or approximate) intermediate value theorem holds if our universe is closed under a closure condition (recursion scheme) which is used in [6] to characterize the polynomial time computable functions. (The results in Sections 4 and 5 are based on the results in [7].)

There are other treatments of relativization in classical framework: see [13, 16] for relativization to the computable functions, and [11] for relativization to some complexity classes.

Within classical logic, we can show that, for any sentence ϕ (say, take the continuum hypothesis in ZFC or a Gödel sentence in a recursive extension of PA), there exists a (constant time) computable function $f : \mathbf{N} \to \mathbf{N}$ which decides whether ϕ holds or $\neg\phi$ holds, that is $(f(0) = 0 \implies \phi) \wedge (f(0) \neq 0 \implies \neg\phi)$. In fact, by virtue of the principle of excluded middle, either ϕ or $\neg\phi$. If ϕ holds, then define $f(x) := 0$; or else $\neg\phi$ holds, define $f(x) := 1$. In either case, such f exists, and hence required f exists.

This classical existence proof of the (constant time) computable function f does *not* ensure that f is actually implementable on a computer. We can not implement f (which is the constant function 0 or the constant function 1 depending on ϕ or $\neg\phi$) on a computer without knowing either ϕ holds or $\neg\phi$ holds.

Here we work within a constructive framework. Because an existence proof of a computable function in constructive (intuitionistic) logic *does* ensure that the function is actually implementable on a computer.

2 Polytime and Elementary Functions

In this section, we review the class $\mathcal{F}$PTIME of polynomial time computable functions and the class $\mathcal{E}$ of elementary functions which is used as a base of the system in [8]. For more details, see [14, 3].

Cobham [4] characterized the class $\mathcal{F}$PTIME of polynomial time computable functions as the following function algebra; see [6] for other characterizations of the polynomial time computable functions by function algebras.

Definition 1 The class of *polynomial time computable function*, $\mathcal{F}$PTIME , is generated by the following clauses

1. $Z, s_0, s_1,\ \#\ , P_i^n$ for $i < n$, $1 \leqslant n$ belong to $\mathcal{F}$PTIME. Here Z is the *zero*-function satisfying $Z(x) = 0$, s_0 and s_1 are the *binary successor* functions satisfying $s_0(x) = 2 \cdot x$ and $s_1(x) = 2 \cdot x + 1$, $\#$ is the *smash* function satisfying $x \# y = 2^{|x|\cdot|y|}$ where $|x|$ denotes the length of the binary representation of x, that is $|x| = \lceil \log(x + 1)\rceil$, and P_i^n is a *projection* determined by $P_i^n(x_0, \dots, x_{n-1}) = x_i$.
2. $\mathcal{F}$PTIME is closed under *composition*: if $f, g_1, \dots, g_m \in \mathcal{F}$PTIME with $f : \mathbf{N}^m \to \mathbf{N}$ and $g_i : \mathbf{N}^n \to \mathbf{N}$ for $1 \leqslant i \leqslant m$, then there is an $h \in \mathcal{F}$PTIME satisfying
$$h(\vec{x}) = f(g_1(\vec{x}), \dots, g_m(\vec{x})).$$
3. $\mathcal{F}$PTIME is closed under the *bounded recursion on notation*: if $f, g_0, g_1, b \in \mathcal{F}$PTIME with $f : \mathbf{N}^n \to \mathbf{N}$, $g_i : \mathbf{N}^{n+2} \to \mathbf{N}$ for $i = 0, 1$ and $b : \mathbf{N}^{n+1} \to \mathbf{N}$, then there is an $h \in \mathcal{F}$PTIME satisfying
$$\begin{aligned} h(0, \vec{y}) &= f(\vec{y}), \\ h(s_0(x), \vec{y}) &= g_0(x, \vec{y}, h(x, \vec{y})) \quad (\text{if } x \neq 0), \\ h(s_1(x), \vec{y}) &= g_1(x, \vec{y}, h(x, \vec{y})), \end{aligned}$$
provided that $h(x, \vec{y}) \leqslant b(x, \vec{y})$ for all $x, \vec{y}$.

The functions $\lfloor x/2 \rfloor$, $\text{MOD}2(x) = x - 2 \cdot \lfloor x/2 \rfloor$, $msp(x, y) = \lfloor x/2^{|y|} \rfloor$, $lsp(x, y) = x \mod 2^{|y|}$, and $pad(x, y) = x \cdot 2^{|y|}$ belong to $\mathcal{F}$PTIME. The coefficient $bit(x, y)$ of $2^{|y|}$ in the binary representation of x is defined by $bit(x, y) := \text{MOD}2(msp(x, y))$. The concatenation function $x * y = x \cdot 2^{|y|} + y$ belongs to $\mathcal{F}$PTIME too.

For a class $\mathcal{U}$ of functions, we let $(\mathcal{U})_*$ denote the class of relations R whose *characteristic function* χ_R, satisfying $R(\vec{x}) \Longleftrightarrow \chi_R(\vec{x}) = 0$, belongs to $\mathcal{U}$. Since $1 \dot{-} x = msp(1, x)$, the signum functions $sg(x) = (1 \dot{-} (1 \dot{-} x))$ and $\overline{sg}(x) = 1 \dot{-} x$ belong to $\mathcal{F}$PTIME. The characteristic functions of $\neg P(\vec{x})$ and $P(\vec{x}) \vee Q(\vec{x})$ can be defined by $\overline{sg}(\chi_P(\vec{x}))$ and $msp(sg(\chi_P(\vec{x})), \overline{sg}(\chi_Q(\vec{x})))$ respectively, and hence the class $(\mathcal{F}\text{PTIME})_*$ is closed under boolean operations.

The conditional function *cond* satisfying

$$cond(x, y, z) = \begin{cases} y & \text{if } x = 0, \\ z & \text{otherwise} \end{cases}$$

belongs to $\mathcal{F}$PTIME. Hence it is easy to see that $\mathcal{F}$PTIME is closed under *definition by cases*: if $f_1, \dots, f_n \in \mathcal{F}$PTIME and $R_1, \dots, R_n$ are disjoint and exhaustive relations in $(\mathcal{F}\text{PTIME})_*$, then there is a $g \in \mathcal{F}$PTIME satisfying

$$g(\vec{x}) = \begin{cases} f_1(\vec{x}) & \text{if } R_1(\vec{x}), \\ & \vdots \\ f_n(\vec{x}) & \text{if } R_n(\vec{x}). \end{cases}$$

Let $\overline{x}$ denote $2^x \dot{-} 1$ whose binary representation consists of x many 1's. Note that $|\overline{x}| = x$. The functions $\overline{|x|}$, $\overline{|x| \dot{-} |y|}$, $\overline{|x| + |y|}$ and $\overline{|x| \cdot |y|}$ belong to $\mathcal{F}$PTIME , and the relations $|x| = |y|$, $|x| < |y|$ and $|x| \leqslant |y|$ belong to $(\mathcal{F}\text{PTIME})_*$.

The *successor* function S with $S(x) = x + 1$, the *predecessor* function prd with $prd(x) = x \dot{-} 1$, the *addition* $x + y$, the *cut-off subtraction* $x \dot{-} y$ and the *multiplication* $x \cdot y$ belong to $\mathcal{F}$PTIME , and the relations $x = y$, $x < y$ and $x \leqslant y$ belong to $(\mathcal{F}\text{PTIME})_*$. The *maximum* and *minimum* functions are defined by $\max\{x, y\} := x + (x \dot{-} y)$ and $\min\{x, y\} := x \dot{-} (x \dot{-} y)$, respectively.

We say that a class $\mathcal{C}$ of relations is closed under *sharply bounded quantifiers* if $(\exists i < |y|)R\left(\overline{i}, \vec{z}\right)$ and $(\forall i < |y|)R\left(\overline{i}, \vec{z}\right)$ are in $\mathcal{C}$ whenever R is in $\mathcal{C}$. It is easy to see that the class $(\mathcal{F}\text{PTIME})_*$ is closed under sharply bounded quantifiers.

The class $\mathcal{F}$PTIME is closed under *sharply bounded minimization*: if $f \in \mathcal{F}$PTIME, then there is a $g \in \mathcal{F}$PTIME satisfying

$$g(y, \vec{z}) = \begin{cases} \overline{\min\left\{i \leqslant |y| \mid f\left(\overline{i}, \vec{z}\right) = 0\right\}} & \text{if } \exists i \leqslant |y| \left(f\left(\overline{i}, \vec{z}\right) = 0\right), \\ \overline{|y| + 1} & \text{otherwise.} \end{cases}$$

We usually write $\overline{\mu i \leqslant |y| \left[f\left(\overline{i}, \vec{z}\right) = 0\right]}$ for g.

Within $\mathcal{F}$PTIME, we can code n-tuples of natural numbers and finite sequence of natural numbers. Let

$$\begin{aligned} \iota(x, y) &:= s_0(x) * \overline{|y|}, \\ \iota_1(z) &:= msp\left(z, \overline{\mu i < |z| \left[bit\left(z, \overline{i}\right) = 0\right] + 1}\right), \\ \iota_2(z) &:= lsp\left(z, \overline{\mu i < |z| \left[bit\left(z, \overline{i}\right) = 0\right]}\right). \end{aligned}$$

Then $\iota_1(\iota(x, y)) = x$ and $\iota_2(\iota(x, y)) = \overline{|y|}$. For the coding of pairs we have a function $j \in \mathbf{N}^2 \to \mathbf{N}$ with inverses j_1, j_2 defined by

$$
\begin{aligned}
j(x, y) &:= \iota(x * y, y), \\
j_1(z) &:= msp(\iota_1(z), \iota_2(z)) \\
j_2(z) &:= lsp(\iota_1(z), \iota_2(z)).
\end{aligned}
$$

Note that, though $j_1(j(x, y)) = x$ and $j_2(j(x, y)) = y$, the function j is not surjective. From j, j_1, j_2 we can construct codings ν^n for n-tuples with inverse j_i^n such that $j_i^n \nu^n(x_1, \ldots, x_n) = x_i \quad (1 \leqslant i \leqslant n)$. We can code finite sequences $x_0, \ldots, x_{|n|-1}$ with $|x_i| \leqslant |b|$ for $i < |n|$ by a number

$$\langle x_0, \ldots, x_{|n|-1} \rangle := \iota\left(\iota\left(\textstyle\sum_{i<n} x_i \cdot 2^{|b|\cdot i}, \overline{|b|}\right), \overline{|n|}\right)$$

with the *length* function $lh(a) := \iota_2(a)$ such that $|lh(a)| = |n|$ and the *decoding* function

$$\pi(a, i) := lsp(msp(\iota_1(\iota_1(a)), \overline{|i| \cdot |\iota_2(\iota_1(a))|}), \overline{|\iota_2(\iota_1(a))|})$$

such that $\pi(a, i) = x_{|i|}$. We usually write $(a)_i$ for $\pi(a, i)$. Note that, letting $\sigma(b, n) := s_0(s_1(b) \# s_1(n)) = 2^{(|b|+1)\cdot(|n|+1)+1}$, we have

$$\langle x_0, \ldots, x_{|n|-1} \rangle < \sigma(b, n)$$

whenever $|x_i| \leqslant |b|$ for all $i < |n|$.

Furthermore, note that if a class $\mathcal{U}$ of functions contains Z, s_0, s_1, P_i^n, msp, lsp, $*$, $\overline{|x|}$, $\overline{|x| + |y|}$ and $\overline{|x| \cdot |y|}$, and is closed under composition and sharply bounded minimization, then we can code n-tuples of natural numbers and finite sequences of natural numbers in $\mathcal{U}$.

The class of elementary functions introduced by Kalmár [10] and Csillag [5] is defined by the following function algebra.

Definition 2 The class of *Kalmár's elementary functions*, $\mathcal{E}$, is generated by the following clauses.

1. Z, S, prd, $+$, $\dot{-}$, $\cdot$, P_i^n for $i < n$, $1 \leqslant n$ belong to $\mathcal{E}$.
2. $\mathcal{E}$ is closed under composition.
3. $\mathcal{E}$ is closed under *bounded sum* and *bounded product*: if $f \in \mathcal{E}$ with $f \in \mathbf{N}^{n+1} \to \mathbf{N}$, then there are $g, h \in \mathcal{E}$ such that

$$
\begin{aligned}
g(0, \vec{x}) &= 0, \qquad g(Sy, \vec{x}) = f(y, \vec{x}) + g(y, \vec{x}) \\
h(0, \vec{x}) &= 1, \qquad h(Sy, \vec{x}) = f(y, \vec{x}) \cdot h(y, \vec{x}).
\end{aligned}
$$

We usually write $\sum_{z<y} f(z, \vec{x})$ and $\prod_{z<y} f(z, \vec{x})$ for g and h, respectively.

It is trivial that the signum functions $sg(x) = 1 \dot{-} (1 \dot{-} x)$ and $\overline{sg}(x) = 1 \dot{-} x$ belong to $\mathcal{E}$. The class $\mathcal{E}$ is closed under *bounded minimum operator*: if $f \in \mathcal{E}$ with $f \in \mathbf{N}^{n+1} \to \mathbf{N}$, then

$$\min_{z \leqslant y}(g(z, \vec{x}) = 0) := \textstyle\sum_{u<Sy} sg\left(\prod_{z<Su} g(z, \vec{x})\right)$$

belongs to $\mathcal{E}$. Note that $\min_{z\leqslant y}(g(z, \vec{x}) = 0)$ is the least $z \leqslant y$ with $f(z, \vec{x}) = 0$, if existing, $y + 1$ otherwise.

It is easy to see that the class $(\mathcal{E})_*$ of elementary relations is closed under boolean operations. Since $\mathcal{E}$ is closed under bounded minimization, $(\mathcal{E})_*$ is closed under *bounded quantifiers*: if R is in $(\mathcal{E})_*$, then $(\exists i < y)R\left(i, \vec{z}\right)$ and $(\forall i < y)R\left(i, \vec{z}\right)$ are in $(\mathcal{E})_*$. The function *cond* is define by $cond(x, y, z) := \overline{sg}(x) \cdot y + sg(x) \cdot z$, and hence $\mathcal{E}$ is closed under definition by cases.

It is clear that the function 2^x belongs to $\mathcal{E}$. The function $\overline{x} = 2^x \dot{-} 1$ belongs to $\mathcal{E}$, and the length in binary function $|x|$ is defined by $|x| := \min_{y\leqslant x}(x < 2^y)$. Hence the functions $x * y = x \cdot 2^{|y|} + y$, $\overline{|x|}$, $\overline{|x| + |y|}$ and $\overline{|x| \cdot |y|}$ belong to $\mathcal{E}$, and $\mathcal{E}$ is closed under sharply bounded minimization. The quotient and remainder functions are defined by $q(x, y) := \min_{z\leqslant x}(x < Sz \cdot y)$ and $r(x, y) := x \dot{-} q(x, y) \cdot y$, respectively. Therefore $msp(x, y) := q(x, 2^{|y|})$ and $lsp(x, y) := r(x, 2^{|y|})$ belong to $\mathcal{E}$.

Within $\mathcal{E}$, we can code n-tuples of natural numbers and finite sequences of natural numbers. Since $\mathcal{E}$ is closed under bounded minimization, we can show the following.

The following theorem is very well known.

Theorem 3 $\mathcal{F}$PTIME $\subseteq \mathcal{E}$.

3 Relativizing to a Universe

As mentioned in [15, 5.4.6], the constructive theory of real numbers may be correct if we relativize the notion of sequence (or unary function) of natural numbers to some universe $\mathcal{U}$ of functions of natural numbers. For a smooth theory of real numbers, $\mathcal{U}$ have to contain some functions and to satisfy certain closure conditions.

We assume that a universe $\mathcal{U}$ of functions of natural numbers contains functions Z, s_0, s_1, P_i^n, msp, lsp, pad, $\overline{|x|}$, $\overline{|x| + |y|}$, $\overline{|x| \cdot |y|}$, $+$ and $\dot{-}$, and is closed under composition and sharply bounded minimization. Then the functions ι, ι_1, ι_2, the pairing function j, and its inverse j_1, j_2 belong to $\mathcal{U}$.

We let a number a code the integer $j_1(a) - j_2(a)$. We denote the set of (code of) integers by $\mathbf{Z}$. Note that $\mathbf{Z} = \mathbf{N}$. The equality $=_\mathbf{Z}$ and the orderings $<_\mathbf{Z}$ and $\leqslant_\mathbf{Z}$ on $\mathbf{Z}$ are defined by $a =_\mathbf{Z} b := j_1(a) + j_2(b) = j_2(a) + j_1(b)$, $a <_\mathbf{Z} b := j_1(a) + j_2(b) <$

$j_2(a) + j_1(b)$, and $a \leqslant_{\mathbf{Z}} b := j_1(a) + j_2(b) \leqslant j_2(a) + j_1(b)$, respectively. The arithmetic operations on $\mathbf{Z}$ are defined by

$$\begin{array}{rcl} a +_{\mathbf{Z}} b & := & j(j_1(a) + j_1(b), j_2(a) + j_2(b)), \\ -_{\mathbf{Z}} a & := & j(j_2(a), j_1(a)), \\ pad_{\mathbf{Z}}(a, n) & := & j(pad(j_1(a), n), pad(j_2(a), n)), \\ \max_{\mathbf{Z}}\{a, b\} & := & j(\max\{j_1(a) + j_2(b), j_2(a) + j_1(b)\}, j_2(a) + j_2(b)), \\ \min_{\mathbf{Z}}\{a, b\} & := & j(\min\{j_1(a) + j_2(b), j_2(a) + j_1(b)\}, j_2(a) + j_2(b)), \\ |a|_{\mathbf{Z}} & := & j((j_1(a) \dot{-} j_2(a)) + (j_2(a) \dot{-} j_1(a)), 0), \end{array}$$

etc. We write $a \cdot 2^{|n|}$ for $pad_{\mathbf{Z}}(a, n)$. The set of natural numbers may be identified with a set of integers by associating with each natural number n the (code of) integer $n^{*\mathbf{Z}} := j(n, 0)$.

We let a number p code the dyadic rational $\iota_1(p)/2^{|\iota_2(p)|}$. We denote the set of (code of) dyadic rationals by $\mathbf{Q}_d$. Note that $\mathbf{Q}_d = \mathbf{N}$. The equality $=_{\mathbf{Q}_d}$ and the orderings $<_{\mathbf{Q}_d}$ and $\leqslant_{\mathbf{Q}_d}$ on $\mathbf{Q}_d$ are defined by $p =_{\mathbf{Q}_d} q := \iota_1(p) \cdot 2^{|\iota_2(q)|} =_{\mathbf{Z}} \iota_1(q) \cdot 2^{|\iota_2(p)|}$, $p <_{\mathbf{Q}_d} q := \iota_1(p) \cdot 2^{|\iota_2(q)|} <_{\mathbf{Z}} \iota_1(q) \cdot 2^{|\iota_2(p)|}$, and $p \leqslant_{\mathbf{Q}_d} q := \iota_1(p) \cdot 2^{|\iota_2(q)|} \leqslant_{\mathbf{Z}} \iota_1(q) \cdot 2^{|\iota_2(p)|}$, respectively. The arithmetic operation on $\mathbf{Q}_d$ are defined by

$$\begin{array}{rcl} p +_{\mathbf{Q}_d} q & := & \iota(\iota_1(p) \cdot 2^{|\iota_2(q)|} +_{\mathbf{Z}} \iota_1(q) \cdot 2^{|\iota_2(p)|}, \overline{|\iota_2(p)| + |\iota_2(q)|}) \\ -_{\mathbf{Q}_d} p & := & \iota(-_{\mathbf{Z}} \iota_1(p), \iota_2(p)), \\ \max_{\mathbf{Q}_d}\{p, q\} & := & \iota(\max_{\mathbf{Z}}\{\iota_1(p) \cdot 2^{|\iota_2(q)|}, \iota_1(q) \cdot 2^{|\iota_2(p)|}\}, \overline{|\iota_2(p)| + |\iota_2(q)|}), \\ \min_{\mathbf{Q}_d}\{p, q\} & := & \iota(\min_{\mathbf{Z}}\{\iota_1(p) \cdot 2^{|\iota_2(q)|}, \iota_1(q) \cdot 2^{|\iota_2(p)|}\}, \overline{|\iota_2(p)| + |\iota_2(q)|}), \\ |p|_{\mathbf{Q}_d} & := & \iota(|\iota_1(p)|_{\mathbf{Z}}, \iota_2(p)), \end{array}$$

etc. The set of integers can be embedded into the set of dyadic rationals by setting $a^{*\mathbf{Q}_d} := \iota(a, 0)$ for each (code of) integer a. We will omit subscripts $\mathbf{Z}$ and $\mathbf{Q}_d$ when there will be no confusion.

Now we can define a notion of a real number relativized to a universe $\mathcal{U}$. A *$\mathcal{U}$-sequence* is a unary function in $\mathcal{U}$.

Definition 4 A *$\mathcal{U}$-real number* is a $\mathcal{U}$-sequence $(p_n)_n$ of dyadic rationals such that

$$\forall mn \in \mathbf{N} \left(|p_m - p_n| < 2^{-|m|} + 2^{-|n|} \right). \tag{†}$$

We shall use a notation $(p_n)_n \in \mathbf{R}(\mathcal{U})$ to mean $(p_n)_n$ is a $\mathcal{U}$-real number. Two $\mathcal{U}$-real numbers $x := (p_n)_n$ and $y := (q_n)_n$ are *equal*, denoted by $x = y$, if

$$\forall n \in \mathbf{N} \left(|p_n - q_n| \leqslant 2^{-|n|+1} \right). \tag{‡}$$

Although the definition of the relativization of real numbers to a universe $\mathcal{U}$ is quite natural and seems quite promising, we cannot say that the relativization of real numbers is really significant and important in constructive analysis without showing that $\mathbf{R}(\mathcal{U})$ and $\mathbf{R}(\mathcal{V})$ actually give two distinct classes of real numbers for some small $\mathcal{U}$ and $\mathcal{V}$. Fortunately, as we will see in the following, $\mathbf{R}(\mathcal{F}\text{PTIME})$ and $\mathbf{R}(\mathcal{E})$ are distinct as sets of real numbers.

Before proving this, we clarify the meaning that $\mathbf{R}(\mathcal{U})$ and $\mathbf{R}(\mathcal{V})$ are identical as sets of real numbers. We say that $\mathbf{R}(\mathcal{U})$ is *a subset of* $\mathbf{R}(\mathcal{V})$ *as a set of real numbers*, denoted by the usual set inclusion $\mathbf{R}(\mathcal{U}) \subseteq \mathbf{R}(\mathcal{V})$, if for each $(p_n)_n \in \mathbf{R}(\mathcal{U})$ there exists $(q_n)_n \in \mathbf{R}(\mathcal{V})$ such that (‡) holds, and that $\mathbf{R}(\mathcal{U})$ and $\mathbf{R}(\mathcal{V})$ are *identical as sets of real numbers*, denoted by $\mathbf{R}(\mathcal{U}) \approx \mathbf{R}(\mathcal{V})$, if $\mathbf{R}(\mathcal{U}) \subseteq \mathbf{R}(\mathcal{V})$ and $\mathbf{R}(\mathcal{V}) \subseteq \mathbf{R}(\mathcal{U})$.

We start to prove that $\mathbf{R}(\mathcal{F}\text{PTIME}) \subsetneq \mathbf{R}(\mathcal{E})$ with quoting the following form of the normal form theorem. (Note that using codings of n-tuples and finite sequences, we can also code a Turing machine (TM) or a randam access machine (RAM) in the usual way.)

Theorem 5 *There exist elementary functions C and D such that $C(e, x, s) \neq 0$ if and only if e is a code of a RAM and s is the number of steps in the computation of this machine at the input x, and if $C(e, x, s) \neq 0$, then $D(C(e, x, s))$ is the output of the computation of the machine given by e at the input x.*

It is known that $C \notin \mathcal{F}\text{PTIME}$ though there exists $C_P \in \mathcal{F}\text{PTIME}$ with $C_P(e, x, \overline{s}) = C(e, x, s)$. We may assume that $C(e, x, s) \neq 0 \implies C(e, x, s) = C(e, x, s+1)$.

Since the class of functions, that can be computed by a RAM running in polynomial time, coincides with $\mathcal{F}\text{PTIME}$ ([14, Theorem 4.4, Chapter 4] or [3, 3.19]), we have the following theorem.

Theorem 6 *Let f be a unary function. Then $f \in \mathcal{F}\text{PTIME}$ if and only if there exist a code e of a RAM and a polynomial p such that $C(e, x, s) \neq 0$ and $f(x) = D(C(e, x, s))$ whenever $p(|x|) \leqslant s$.*

The following is an auxiliary lemma.

Lemma 7 *Let $Q(a, n)$ be a relation defined by*

$$Q(a, n) := \left(a = \sum_{i<|n|} 3 \cdot bit\left(a, \overline{2i}\right) \cdot 2^{2i}\right).$$

Then Q belongs to $(\mathcal{F}\text{PTIME})_*$, and $Q(a, n) \wedge Q(b, n) \wedge a \neq b \implies 3 \leqslant |a - b|$ for each $n, a, b \in \mathbf{N}$,

Proof Since the function $h(a, n) = \sum_{i<|n|} 3 \cdot bit\left(a, \overline{2i}\right) \cdot 2^{2i}$ is defined by

$$\begin{array}{rcll} h(a, 0) & = & 0, & \\ h(a, s_0(n)) & = & h(a, n) + 3 \cdot bit(a, \overline{2|n|}) \cdot 2^{2|n|} & (\text{if } n \neq 0), \\ h(a, s_1(n)) & = & h(a, n) + 3 \cdot bit(a, \overline{2|n|}) \cdot 2^{2|n|} & \end{array}$$

and $h(a, n) \leqslant n \# 2$, the relation $Q(a, n) := (a = h(a, n))$ belongs to $(\mathcal{F}\text{PTIME})_*$. Suppose that $Q(a, n)$, $Q(b, n)$ and $a < b$. Then, taking the maximum i with $i < |n|$ and $bit\left(a, \overline{2i}\right) \neq bit\left(b, \overline{2i}\right)$, we have $a = u \cdot 2^{i+2} + 0 \cdot 2^i + v$ and $b = u \cdot 2^{i+2} + 3 \cdot 2^i + w$ for some u, v, w with $i = |v| = |w|$, and hence

$$\text{b - a} = 3 \cdot 2^i + (w - v) \geqslant 3 \cdot 2^i - (2^i \dot{-} 1) \geqslant 3. \ \square$$

In the following Proposition, we construct an elementary function, which is not in $\mathcal{F}$PTIME, with special properties.

Proposition 8 *There exists a function* $\alpha \in \mathcal{E} \setminus \mathcal{F}$PTIME *such that for every* $m, n \in \mathbf{N}$

1. $Q(\alpha(n), n)$,
2. $\alpha(n) = \alpha(\overline{|n|})$,
3. $\lfloor \alpha(\overline{|n| + |m|})/2^{2|m|} \rfloor = \alpha(\overline{|n|})$.

Proof Let δ be the elementary function defined by

$$\delta(x) := \begin{cases} 1 & \text{if } C(\iota_1(|x|), 2^{|x|}, 2^{|x|}) \neq 0 \text{ and } \mathrm{MOD2}(D(C(\iota_1(|x|, 2^{|x|}, 2^{|x|}))) = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Define a function α by

$$\alpha(n) := \sum_{i<|n|} 3 \cdot \delta(\overline{|n| \dot{-} (i+1)}) \cdot 2^{2i}.$$

Then α belongs to $\mathcal{E}$, and, trivially, $\alpha(n) = \alpha(\overline{|n|})$ for each $n \in \mathbf{N}$. Since $\sum_{i<|n|} 3 \cdot bit(\alpha(n), \overline{2i}) \cdot 2^{2i} = \sum_{i<|n|} 3 \cdot \delta(\overline{|n| \dot{-} (i+1)}) \cdot 2^{2i} = \alpha(n)$, we have $Q(\alpha(n), n)$ for each $n \in \mathbf{N}$. Furthermore, we have

$$\begin{aligned} \lfloor \alpha(\overline{|n| + |m|})/2^{2|m|} \rfloor &= \lfloor \sum_{i<|n|+|m|} 3 \cdot \delta(\overline{|n| + |m| \dot{-} (i+1)}) \cdot 2^{2i}/2^{2|m|} \rfloor \\ &= \sum_{i<|n|} 3 \cdot \delta(\overline{|n| \dot{-} (i+1)}) \cdot 2^{2i} = \alpha(\overline{|n|}). \end{aligned}$$

Suppose that $\alpha \in \mathcal{F}$PTIME . Then there exist a code e of a RAM and a polynomial p such that if $p(|x|) \leqslant s$, then $C(e, x, s) \neq 0$ and $\delta(x) = D(C(e, x, s))$. Choose m so that $p(|2^z|) = p(z+1) \leqslant 2^z$ for $z := \iota(e, m)$. Then, since $C(\iota_1(z), 2^z, 2^z) = C(e, 2^z, 2^z) \neq 0$, we have

$$\begin{aligned} \mathrm{MOD2}(\alpha(2^z)) = 1 &\iff \delta(\overline{|2^z| \dot{-} 1}) = 1 \\ &\iff \delta(\overline{z}) = 1 \\ &\iff \mathrm{MOD2}(D(C(\iota_1(z), 2^z, 2^z))) = 0 \\ &\iff \mathrm{MOD2}(D(C(e, 2^z, 2^z))) = 0 \\ &\iff \mathrm{MOD2}(\alpha(2^z)) = 0, \end{aligned}$$

a contradiction. Thus $\alpha \notin \mathcal{F}$PTIME. $\square$

Note that the function which gives the nearest natural number $[p]$ to a dyadic rational p belongs to $\mathcal{F}$PTIME. In fact, since the nearest natural number $h(b,n)$ to $b/2^{|n|}$ is defined by

$$h(b,n) := \begin{cases} S(msp(b,n)) & \text{if } 2^{|n| \dot{-} 1} \leqslant lsp(b,n), \\ msp(b,n) & \text{otherwise}, \end{cases}$$

we set $[p] := h(j_1(\iota_1(p)) \dot{-} j_2(\iota_1(p)), \iota_2(p))$.

Theorem 9 $\mathbf{R}(\mathcal{F}\text{PTIME}) \subsetneq \mathbf{R}(\mathcal{E})$.

Proof It is trivial that $\mathbf{R}(\mathcal{F}\text{PTIME}) \subseteq \mathbf{R}(\mathcal{E})$. Suppose that $\mathbf{R}(\mathcal{E}) \subseteq \mathbf{R}(\mathcal{F}\text{PTIME})$. Let α be the function constructed in Proposition 8, and define a sequence $x := (p_n)_n$ of dyadic rationals by

$$p_n := \frac{\alpha(\overline{|n|+1})}{2^{2|n|}}.$$

Then, since if $|m| \leqslant |n|$ then

$$\begin{aligned} |p_n - p_m| &= 2^{-2|n|} \cdot \left|\alpha(\overline{|n|+1}) - \alpha(\overline{|m|+1}) \cdot 2^{2(|n| \dot{-} |m|)}\right| \\ &= 2^{-2|n|} \cdot \left|\alpha(\overline{|n|+1}) - \lfloor \alpha(\overline{|n|+1})/2^{2(|n| \dot{-} |m|)} \rfloor \cdot 2^{2(|n| \dot{-} |m|)}\right| \\ &< 2^{-2|n|} \cdot 2^{2(|n| \dot{-} |m|)} = 2^{-2|m|} < 2^{-|n|} + 2^{-|m|}, \end{aligned}$$

by Proposition 8 (3) and the fact that $a - \lfloor a/2^{|b|} \rfloor \cdot 2^{|b|} < 2^{|b|}$ for each $a, b \in \mathbf{N}$, we have $x \in \mathbf{R}(\mathcal{E})$. Hence there exists $y := (q_n)_n \in \mathbf{R}(\mathcal{F}\text{PTIME})$ such that

$$\forall n \in \mathbf{N} \left(|p_n - q_n| \leqslant 2^{-|n|+1}\right).$$

Let $\beta(n) := [2^{2|n|-2}|q_{\overline{2|n|}}|]$. Then $\beta \in \mathcal{F}$PTIME, and, since $0 \leqslant 2^{2|n|-2}|q_{\overline{2|n|}}|$, we have

$$\left|2^{2|n|-2}|q_{\overline{2|n|}}| - \beta(n)\right| = \left|2^{2|n|-2}|q_{\overline{2|n|}}| - [2^{2|n|-2}|q_{\overline{2|n|}}|]\right| \leqslant 2^{-1}.$$

Hence

$$\begin{aligned} \left|2^{2|n|-2} p_{\overline{2|n|}} - \beta(n)\right| &\leqslant \left|2^{2|n|-2} p_{\overline{2|n|}} - 2^{2|n|-2}|q_{\overline{2|n|}}|\right| + \left|2^{2|n|-2}|q_{\overline{2|n|}}| - \beta(n)\right| \\ &\leqslant 2^{2|n|-2} \cdot \left|p_{\overline{2|n|}} - |q_{\overline{2|n|}}|\right| + 2^{-1} \\ &= 2^{2|n|-2} \cdot \left||p_{\overline{2|n|}}| - |q_{\overline{2|n|}}|\right| + 2^{-1} \\ &\leqslant 2^{2|n|-2} \cdot \left|p_{\overline{2|n|}} - q_{\overline{2|n|}}\right| + 2^{-1} \leqslant 2^{2|n|-2} \cdot 2^{-2|n|+1} + 2^{-1} \\ &= 2^{-1} + 2^{-1} = 1 \end{aligned}$$

for each $n \in \mathbf{N}$. Since $\lfloor 2^{2|n|-2} p_{\overline{2|n|}} \rfloor = \lfloor \alpha(\overline{2|n|+1})/2^{2(|n|+1)} \rfloor = \alpha(\overline{|n|})$, by Proposition 8 (3), we have, by Proposition 8 (2),

$$\begin{aligned} |\alpha(n) - \beta(n)| &\leqslant \left|\alpha(\overline{|n|}) - 2^{2|n|-2} p_{\overline{2|n|}}\right| + \left|2^{2|n|-2} p_{\overline{2|n|}} - \beta(n)\right| \\ &\leqslant \left|\lfloor 2^{2|n|-2} p_{\overline{2|n|}} \rfloor - 2^{2|n|-2} p_{\overline{2|n|}}\right| + 1 \\ &< 1 + 1 = 2. \end{aligned}$$

Therefore, since $Q(\alpha(n), n)$, by Proposition 8 (1), and $Q(a, n) \wedge Q(b, n) \wedge a \neq b \implies 3 \leqslant |a - b|$ for each $a, b, n \in \mathbf{N}$, by Lemma 7, we see that $\alpha(n)$ is the unique element a in $\{\beta(n) \dot{-} 1, \beta(n), \beta(n) + 1\}$ satisfying $Q(a, n)$, and so, we can define α by

$$\alpha(n) := \begin{cases} \beta(n) \dot{-} 1 & \text{if } Q(\beta(n) \dot{-} 1, n) \\ \beta(n) & \text{if } Q(\beta(n), n) \\ \beta(n) + 1 & \text{if } Q(\beta(n) + 1, n). \end{cases}$$

Thus $\alpha \in \mathcal{F}$PTIME, a contradiction. □

4 $\mathcal{U}$-Real Numbers

We have seen that the relativization of real numbers to a universe $\mathcal{U}$ is meaningful. In this section, we develop a basic theory of relativized real numbers, and show the Cauchy completeness of relativized reals $\mathbf{R}(\mathcal{U})$.

Let $x := (p_n)_n$, $y := (q_n)_n \in \mathbf{R}(\mathcal{U})$. Then the ordering relation $<$ between x and y is defined by

$$x < y := \exists n \in \mathbf{N} \left(2^{-|n|+1} < q_n - p_n\right). \tag{$*$}$$

We define $x \leqslant y$ to mean $\neg(y < x)$. The inequality $\neq$ is given by $x \neq y := (x < y \vee y < x)$. It is readily to see that $x = y \iff x \leqslant y \wedge y \leqslant x \iff \neg(x \neq y)$.

Lemma 10 *Let* $x, y, z \in \mathbf{R}(\mathcal{U})$. *Then*

1. $\neg(x < y \wedge y < x)$,
2. $x < y \implies x < z \vee z < y$.

Proof (1). Let $x = (p_n)_n$ and $y = (q_n)_n$, and suppose that $x < y$ and $y < x$. Then, by ($*$), there exist n, n' such that $2^{-|n|+1} < q_n - p_n$ and $2^{-|n'|+1} < p_{n'} - q_{n'}$, and hence, by (†),

$$\begin{aligned} 0 &= (q_n - p_n) + (p_{n'} - q_{n'}) - (p_{n'} - p_n) - (q_n - q_{n'}) \\ &> 2^{-|n|+1} + 2^{-|n'|+1} - (2^{-|n'|} + 2^{-|n|}) - (2^{-|n|} + 2^{-|n'|}) = 0 \end{aligned}$$

a contradiction.

(2). Let $x = (p_n)_n$, $y = (q_n)_n$ and $z = (r_n)_n$, and suppose that $x < y$. Then, by (∗), there exists n such that $2^{-|n|+1} < q_n - p_n$, and hence there exists N such that $2^{-|n|+1} + 2^{-|N|+3} < q_n - p_n$. Either $(p_n + q_n)/2 < r_N$ or $r_N \leqslant (p_n + q_n)/2$. In the former case, we have

$$\begin{aligned} r_N - p_N &> \frac{p_n + q_n}{2} - p_N = \frac{p_n + q_n}{2} - p_n - (p_N - p_n) \\ &= \frac{q_n - p_n}{2} - (p_N - p_n) > 2^{-|n|} + 2^{-|N|+2} - (2^{-|N|} + 2^{-|n|}) \\ &= 3 \cdot 2^{-|N|} > 2^{-|N|+1}, \end{aligned}$$

by (†), and hence $x < z$. In the latter case, we have

$$\begin{aligned} q_N - r_N &\geqslant q_N - \frac{p_n + q_n}{2} = (q_N - q_n) + q_n - \frac{p_n + q_n}{2} \\ &= (q_N - q_n) + \frac{q_n - p_n}{2} > -(2^{-|N|} + 2^{-|n|}) + 2^{-|n|} + 2^{-|N|+2} \\ &> 2^{-|N|+1}, \end{aligned}$$

by (†), and hence $z < y$. □

Proposition 11 *Let $x, x', y, y', z \in \mathbf{R}(\mathcal{U})$. Then*

1. $x = x' \wedge y = y' \wedge x < y \implies x' < y'$,
2. $x = x' \wedge y = y' \wedge x \leqslant y \implies x' \leqslant y'$,
3. $x = x' \wedge y = y' \wedge x \neq y \implies x' \neq y'$,
4. $x < y \wedge y < z \implies x < z$,
5. $x < y \wedge y \leqslant z \implies x < z$,
6. $x \leqslant y \wedge y < z \implies x < z$,
7. $x \leqslant y \wedge y \leqslant z \implies x \leqslant z$,
8. $x \neq y \implies x \neq z \vee z \neq y$.

Proof Here we show (1) and (4), and the rest are left to the reader.

(1) Suppose that $x = x'$, $y = y'$ and $x < y$. Then either $x < x'$ or $x' < y$ by Lemma 10(2). Since $\neg(x < x')$, the latter must be the case. Hence $x' < y'$ or $y' < y$, by Lemma 10(2), and therefore, since $\neg(y' < y)$, we have $x' < y'$.

(4) Suppose that $x < y$ and $y < z$ Then either $x < z$ or $z < y$, by Lemma 10(2). In the latter case, we have a contradiction by Lemma 10(1). Thus the former must be the case.

The arithmetic operations on $\mathbf{R}(\mathcal{U})$ are defined in terms of operations on sequences of dyadic rationals as follows:

$$\begin{aligned} x + y &:= (p_{\overline{|n|+1}} + q_{\overline{|n|+1}})_n, \\ -x &:= (-p_n)_n, \\ \max\{x, y\} &:= (\max\{p_n, q_n\})_n, \\ \min\{x, y\} &:= (\min\{p_n, q_n\})_n, \\ |x| &:= (|p_n|)_n. \end{aligned}$$

Note that $\min\{x, y\} = -\max\{-x, -y\}$ and $|x| = \max\{x, -x\}$. The set of dyadic rationals may be identified with a set of $\mathcal{U}$-real numbers by associating each dyadic rational p the constant $\mathcal{U}$-sequence $(p, p, p, \ldots)$.

There is no trouble in proving that if $x, y \in \mathbf{R}(\mathcal{U})$, then $x + y, -x, \max\{x, y\}, \min\{x, y\}, |x| \in \mathbf{R}(\mathcal{U})$, and that $(x, y) \mapsto x + y$, $x \mapsto -x$, $(x, y) \mapsto \max\{x, y\}$, $(x, y) \mapsto \min\{x, y\}$ and $x \mapsto |x|$ are functions in the sense that $x = x' \wedge y = y' \implies x + y = x' + y'$, $x = x' \implies -x = -x'$, etc.

Proposition 12 *Let $x, y, z \in \mathbf{R}(\mathcal{U})$. Then*

1. $(x + y) + z = x + (y + z)$,
2. $x + y = y + x$,
3. $x + 0 = x$,
4. $x + (-x) = 0$,
5. $x < y \implies x + z < y + z$,
6. $x \leqslant y \implies x + z \leqslant y + z$,
7. $x \leqslant \max\{x, y\}$, $y \leqslant \max\{x, y\}$,
8. $z < \max\{x, y\} \implies z < x \vee z < y$,
9. $\min\{x, y\} \leqslant x$, $\min\{x, y\} \leqslant y$,
10. $\min\{x, y\} < z \implies x < z \vee y < z$,
11. $0 \leqslant |x|$,
12. $|x + y| \leqslant |x| + |y|$.

Proof Here we show only (5), and the rest are left to the reader.

(5) Let $x = (p_n)_n$, $y = (q_n)_n$ and $z = (r_n)_n$, and suppose that $x < y$. Then, by $(*)$, there exists n such that $2^{-|n|+1} < q_n - p_n$, and hence there exists N such that $2^{-|n|+1} + 2^{-|N|+2} < q_n - p_n$. Therefore

$$\begin{aligned} &(q_{\overline{|N|+1}} + r_{\overline{|N|+1}}) - (p_{\overline{|N|+1}} + r_{\overline{|N|+1}}) \\ &\quad = q_{\overline{|N|+1}} - p_{\overline{|N|+1}} = (q_{\overline{|N|+1}} - q_n) + (q_n - p_n) + (p_n - p_{\overline{|N|+1}}) \\ &\quad > -2^{-|N|-1} - 2^{-|n|} + 2^{-|n|+1} + 2^{-|N|+2} - 2^{-|n|} - 2^{-|N|-1} \\ &\quad = 2^{-|N|+2} - 2^{-|N|} > 2^{-|N|+1}. \end{aligned}$$

Thus $x + z < y + z$, by the definitions of $+$ and $<$. □

Lemma 13 *For each $x := (p_n)_n \in \mathbf{R}(\mathcal{U})$, we have*

$$\forall n \in \mathbf{N} \left(|p_n - x| \leqslant 2^{-|n|}\right).$$

Proof Suppose that $|p_n - x| > 2^{-|n|}$. Then, noting that $|p_n - x|$ is the sequence $(|p_n - p_{\overline{|m|+1}}|)_m$, by the definitions of $+$, $-$ and $|\cdot|$, we can find a number m such that

$$2^{-|m|+1} < \left|p_n - p_{\overline{|m|+1}}\right| - 2^{-|n|} < 2^{-|n|} + 2^{-|m|-1} - 2^{-|n|} = 2^{-|m|-1},$$

a contradiction. □

A *sequence of $\mathcal{U}$-real numbers* $(x_m)_m$ is a double $\mathcal{U}$-sequence $((p_n^m)_n)_m$ of dyadic rationals such that $(p_n^m)_n \in \mathbf{R}(\mathcal{U})$ for each m. A sequence $(x_n)_n$ of $\mathcal{U}$-reals *converges* to $x \in \mathbf{R}(\mathcal{U})$ if there exists a *modulus* $\beta \in \mathcal{U}$ such that

$$\forall kn \in \mathbf{N}\left(|x - x_{\beta(k)+n}| < 2^{-|k|}\right).$$

Then x is said to be the *limit* of $(x_n)_n$. A sequence $(x_n)_n$ of $\mathcal{U}$-reals is a *Cauchy sequence* if there exists a *Cauchy modulus* $\alpha \in \mathcal{U}$ such that

$$\forall kmn \in \mathbf{N}\left(|x_{\alpha(k)+m} - x_{\alpha(k)+n}| < 2^{-|k|}\right).$$

Theorem 14 *Each Cauchy sequence of $\mathcal{U}$-reals converges to a limit.*

Proof Let $(x_m)_m := ((p_n^m)_n)_m$ be a Cauchy sequence of $\mathcal{U}$-reals with Cauchy modulus $\alpha \in \mathcal{U}$, i.e.

$$\forall kmn \in \mathbf{N}\left(|x_{\alpha(k)+m} - x_{\alpha(k)+n}| < 2^{-|k|}\right),$$

and define a $\mathcal{U}$-sequence $(q_n)_n$ of dyadic rationals by

$$q_n := p_{\overline{|n|+1}}^{\alpha(\overline{|n|+1})}.$$

Then $\left|q_n - x_{\alpha(\overline{|n|+1})}\right| \leqslant 2^{-|n|-1}$ for each n, by Lemma 13. If $\alpha(\overline{|m|+1}) \leqslant \alpha(\overline{|n|+1})$, then $\left|x_{\alpha(\overline{|m|+1})} - x_{\alpha(\overline{|n|+1})}\right| < 2^{-|m|-1} < 2^{-|m|-1} + 2^{-|n|-1}$; or else $\alpha(\overline{|n|+1}) < \alpha(\overline{|m|+1})$, and $\left|x_{\alpha(\overline{|m|+1})} - x_{\alpha(\overline{|n|+1})}\right| < 2^{-|n|-1} < 2^{-|m|-1} + 2^{-|n|-1}$. Hence we have

$$\begin{aligned}
|q_m - q_n| &\leqslant \left|q_m - x_{\alpha(\overline{|m|+1})}\right| + \left|x_{\alpha(\overline{|m|+1})} - x_{\alpha(\overline{|n|+1})}\right| + \left|x_{\alpha(\overline{|n|+1})} - q_n\right| \\
&\leqslant 2^{-|m|-1} + 2^{-|m|-1} + 2^{-|n|-1} + 2^{-|n|-1} = 2^{-|m|} + 2^{-|n|}.
\end{aligned}$$

Therefore $x := (q_n)_n \in \mathbf{R}(\mathcal{U})$. Furthermore we have

$$\begin{aligned}
&\left|x - x_{\alpha(\overline{|k|+2})+m}\right| \\
&\quad \leqslant \left|x - q_{\overline{|k|+1}}\right| + \left|q_{\overline{|k|+1}} - x_{\alpha(\overline{|k|+2})}\right| + \left|x_{\alpha(\overline{|k|+2})} - x_{\alpha(\overline{|k|+2})+m}\right| \\
&\quad < 2^{-|k|-1} + 2^{-|k|-2} + 2^{-|k|-2} = 2^{-|k|},
\end{aligned}$$

and hence $(x_n)_n$ converges to x with a modulus $\beta(n) := \alpha(\overline{|n|+2})$. □

5 Intermediate Value Theorem

In this section, we show that the (constructive or approximate) intermediate value theorem holds if our universe is closed under a closure condition (recursion scheme) which is used to characterize the polynomial time computable functions.

Let $[0, 1]_{\mathbf{Q}_\mathrm{d}} := \{r \in \mathbf{Q}_\mathrm{d} \mid 0 \leqslant r \leqslant 1\}$, and $[0, 1]_{\mathbf{R}(\mathcal{U})} := \{x \in \mathbf{R}(\mathcal{U}) \mid 0 \leqslant x \leqslant 1\}$. Then a *uniformly continuous $\mathcal{U}$-function f* from $[0, 1]_{\mathbf{R}(\mathcal{U})}$ to $\mathbf{R}(\mathcal{U})$ consists of two functions φ and μ in $\mathcal{U}$ such that

1. $f(p) := (\varphi(p, n))_n \in \mathbf{R}(\mathcal{U})$ for each $p \in [0, 1]_{\mathbf{Q}_\mathrm{d}}$,
2. for all $k \in \mathbf{N}$ and $p, q \in [0, 1]_{\mathbf{Q}_\mathrm{d}}$,

$$|p - q| < 2^{-|\mu(k)|} \implies |f(p) - f(q)| < 2^{-|k|}.$$

We may assume, without loss of generality, that $|n| \leqslant |\mu(n)|$ for each n.

Proposition 15 *Let f be a uniformly continuous $\mathcal{U}$-function from $[0, 1]_{\mathbf{R}(\mathcal{U})}$ to $\mathbf{R}(\mathcal{U})$ given by φ and μ, let $\alpha(n) := \overline{|\mu(\overline{|n| + 1})|} + 1$, and, for $x := (p_n)_n \in [0, 1]_{\mathbf{R}(\mathcal{U})}$, let*

$$f(x) := \left(\varphi(\hat{p}_{\alpha(n)}, \overline{|n| + 1})\right)_n ,$$

where $\hat{p}_n := \min\{\max\{p_n, 0\}, 1\}$. Then

1. *$f(x) \in \mathbf{R}(\mathcal{U})$ for each $x \in [0, 1]_{\mathbf{R}(\mathcal{U})}$,*
2. *for all $k \in \mathbf{N}$ and $x, y \in [0, 1]_{\mathbf{R}(\mathcal{U})}$,*

$$|x - y| < 2^{-|\alpha(k)|} \implies |f(x) - f(y)| < 2^{-|k|};$$

especially, if $x = y$, then $f(x) = f(y)$.

Proof Note that, since $|\hat{p}_m - \hat{p}_n| \leqslant |p_m - p_n|$ and $|p_n - \hat{p}_n| \leqslant 2^{-|n|+1}$ for each m and n, we have $(\hat{p}_n)_n \in \mathbf{R}(\mathcal{U})$ and $x = (\hat{p}_n)_n$. Since $|\hat{p}_{\alpha(m)} - \hat{p}_{\alpha(n)}| < 2^{-|\alpha(m)|} + 2^{-|\alpha(n)|}$, if $|\alpha(m)| \leqslant |\alpha(n)|$, then $|\hat{p}_{\alpha(m)} - \hat{p}_{\alpha(n)}| < 2^{-|\alpha(m)|+1} = 2^{|\mu(\overline{|m|+1})|}$, and hence $|f(\hat{p}_{\alpha(m)}) - f(\hat{p}_{\alpha(n)})| < 2^{-|m|-1} < 2^{-|m|-1} + 2^{-|n|-1}$; or else $|\alpha(n)| < |\alpha(n)|$, and $|\hat{p}_{\alpha(m)} - \hat{p}_{\alpha(n)}| < 2^{-|\mu(\overline{|n|+1})|}$, and hence $|f(\hat{p}_{\alpha(m)}) - f(\hat{p}_{\alpha(n)})| < 2^{-|m|-1} + 2^{-|n|-1}$. Therefore

$$\begin{aligned}
&\left|\varphi(\hat{p}_{\alpha(m)}, \overline{|m| + 1}) - \varphi(\hat{p}_{\alpha(n)}, \overline{|n| + 1})\right| \\
&\quad \leqslant \quad \left|\varphi(\hat{p}_{\alpha(m)}, \overline{|m| + 1}) - f(\hat{p}_{\alpha(m)})\right| + \left|f(\hat{p}_{\alpha(m)}) - f(\hat{p}_{\alpha(n)})\right| \\
&\qquad\qquad + \left|f(\hat{p}_{\alpha(n)}) - \varphi(\hat{p}_{\alpha(n)}, \overline{|n| + 1})\right| \\
&\quad < \quad 2^{-|m|-1} + 2^{-|m|-1} + 2^{-|n|-1} + 2^{-|n|-1} = 2^{-|m|} + 2^{-|n|},
\end{aligned}$$

for each m and n, and so $f(x)$ is a $\mathcal{U}$-real number.

Let $x := (p_n)_n$, $y := (q_n)_n \in [0, 1]_{\mathbf{R}(\mathcal{U})}$, and suppose that $|x - y| < 2^{-|\alpha(k)|}$. Then, taking $n := \overline{|\alpha(k)| + |k| + 3}$, since $|\alpha(k)| + 1 < |n| + 1 \leqslant \left|\mu(\overline{|n| + 1})\right| < |\alpha(n)|$, we have $|x - \hat{p}_{\alpha(n)}| \leqslant 2^{-|\alpha(n)|} < 2^{-|\alpha(k)|-1}$ and, similarly, $|\hat{q}_{\alpha(n)} - y| < 2^{-|\alpha(k)|-1}$. Hence

$$\begin{array}{rcl} |\hat{p}_{\alpha(n)} - \hat{q}_{\alpha(n)}| & \leqslant & |\hat{p}_{\alpha(n)} - x| + |x - y| + |y - \hat{q}_{\alpha(n)}| \\ & < & 2^{-|\alpha(k)|-1} + 2^{-|\alpha(k)|} + 2^{-|\alpha(k)|-1} = 2^{-|\alpha(k)|+1} = 2^{-\left|\mu\left(\overline{|k|+1}\right)\right|}, \end{array}$$

and therefore $|f(\hat{p}_{\alpha(n)}) - f(\hat{q}_{\alpha(n)})| < 2^{-|k|-1}$. Since

$$\begin{array}{rl} & |f(x) - f(\hat{p}_{\alpha(n)})| \\ \leqslant & \left|f(x) - \varphi(\hat{p}_{\alpha(n)}, \overline{|n|+1})\right| + \left|\varphi(\hat{p}_{\alpha(n)}, \overline{|n|+1}) - f(\hat{p}_{\alpha(n)})\right| \\ \leqslant & 2^{-|n|} + 2^{-|n|-1} < 2^{-|k|-3} + 2^{-|k|-4} < 2^{-|k|-2}, \end{array}$$

and, similarly, $|f(\hat{q}_{\alpha(n)}) - f(y)| < 2^{-|k|-2}$, we have

$$\begin{array}{rl} & |f(x) - f(y)| \\ \leqslant & |f(x) - f(\hat{p}_{\alpha(n)})| + |f(\hat{p}_{\alpha(n)}) - f(\hat{q}_{\alpha(n)})| + |f(\hat{q}_{\alpha(n)}) - f(y)| \\ < & 2^{-|k|-2} + 2^{-|k|-1} + 2^{-|k|-2} = 2^{-|k|}. \end{array}$$

□

In the rest of this section, we assume that our universe $\mathcal{U}$ is closed under *full concatenation recursion on notation* (FCRN) which is used in [6] to characterize the polynomial time computable functions: if $f, g_0, g_1 \in \mathcal{U}$ with $g_0(m, \vec{n}, l), g_1(m, \vec{n}, l) \leqslant 1$, then there is an $h \in \mathcal{U}$ such that

$$\begin{array}{rcll} h(0, \vec{n}) & = & f(\vec{n}), & \\ h(s_0(m), \vec{n}) & = & s_{g_0(m,\vec{n},h(m,\vec{n}))}(h(m, \vec{n})) & (\text{if } m \neq 0), \\ h(s_1(m), \vec{n}) & = & s_{g_1(m,\vec{n},h(m,\vec{n}))}(h(m, \vec{n})). & \end{array}$$

Theorem 16 *Let f be a uniformly continuous $\mathcal{U}$-function from $[0, 1]_{\mathbf{R}(\mathcal{U})}$ to $\mathbf{R}(\mathcal{U})$ with $f(0) \leqslant 0 \leqslant f(1)$. Then*

$$\forall k \in \mathbf{N} \exists x \in [0, 1]_{\mathbf{R}(\mathcal{U})} \left(|f(x)| \leqslant 2^{-|k|}\right).$$

Proof Let f be given by functions φ and μ in $\mathcal{U}$. Let λ be a function defined by

$$\lambda(n, k, m) := \begin{cases} 0 & \text{if } 0 \leqslant \varphi((2m+1)/2^{|n|+1}, \overline{|k|+2}), \\ 1 & \text{otherwise,} \end{cases}$$

and define a function θ by

$$\begin{array}{rcll} \theta(0, k) & = & 0 & \\ \theta(s_0(n), k) & = & s_{\lambda(n,k,\theta(n,k))}(\theta(n, k)), & (\text{if } n \neq 0) \\ \theta(s_1(n), k) & = & s_{\lambda(n,k,\theta(n,k))}(\theta(n, k)). & \end{array}$$

Then, since $\lfloor \theta(\overline{|n|+1}, k)/2 \rfloor = \theta(\overline{|n|}, k)$ for each $n \in \mathbf{N}$, it is straightforward to show, by induction, that $\lfloor \theta(\overline{|n|+|m|}, k)/2^{|m|} \rfloor = \theta(\overline{|n|}, k)$ for each $m, n \in \mathbf{N}$. Let

$$p_{n,k} := \frac{\theta(\overline{|n|}, k)}{2^{|n|}} \quad \text{and} \quad q_{n,k} := \frac{\theta(\overline{|n|}, k) + 1}{2^{|n|}}.$$

We show that

$$\varphi(p_{n,k}, \overline{|k|+2}) \leqslant 2^{-|k|-1} \quad \text{and} \quad -2^{-|k|-1} \leqslant \varphi(q_{n,k}, \overline{|k|+2})$$

for each $n \in \mathbf{N}$. We proceed by induction on $|n|$. If $|n| = 0$, then $p_{n,k} = 0$ and $q_{n,k} = 1$, and hence

$$\varphi(p_{n,k}, \overline{|k|+2}) \leqslant f(0) + \left| f(0) - \varphi(0, \overline{|k|+2}) \right| \leqslant 0 + 2^{-|k|-1} = 2^{-|k|-1}$$

and

$$\varphi(q_{n,k}, \overline{|k|+2}) \geqslant f(1) - \left| f(1) - \varphi(1, \overline{|k|+2}) \right| \geqslant 0 - 2^{-|k|-1} = -2^{-|k|-1}.$$

Suppose that the inequalities hold for n with $|n| = |m|$, and consider the case that $|n| = |m| + 1$. Then $\overline{|n|} = s_1(\overline{|m|})$, and either $\lambda(\overline{|m|}, k, \theta(\overline{|m|}, k)) = 0$ or $\lambda(\overline{|m|}, k, \theta(\overline{|m|}, k)) = 1$. In the former case, since

$$p_{n,k} = \frac{s_0(\theta(\overline{|m|}, k))}{2^{|n|}} = \frac{2\theta(\overline{|m|}, k)}{2^{|m|+1}} = p_{m,k},$$

we have $\varphi(p_{n,k}, \overline{|k|+2}) = \varphi(p_{m,k}, \overline{|k|+2}) \leqslant 2^{-|k|-1}$ by the induction hypothesis, and, since $\lambda(\overline{|m|}, k, \theta(\overline{|m|}, k)) = 0$, we have

$$-2^{-|k|-1} < 0 \leqslant \varphi((2\theta(\overline{|m|}, k) + 1)/2^{|m|+1}, \overline{|k|+2}) = \varphi(q_{n,k}, \overline{|k|+2}).$$

In the latter case, since

$$q_{n,k} = \frac{s_1(\theta(\overline{|m|}, k)) + 1}{2^{|n|}} = \frac{2\theta(\overline{|m|}, k) + 2}{2^{|m|+1}} = q_{m,k},$$

we have $-2^{-|k|-1} \leqslant \varphi(q_{m,k}, \overline{|k|+2}) = \varphi(q_{n,k}, \overline{|k|+2})$ by the induction hypothesis, and, since $\lambda(\overline{|m|}, k, \theta(\overline{|m|}, k)) = 1$, we have

$$\varphi(p_{n,k}, \overline{|k|+2}) = \varphi((2\theta(\overline{|m|}, k) + 1)/2^{2|m|+1}, \overline{|k|+2}) < 0 < 2^{-|k|-1}.$$

Let $x := (p_{n,k})_n$ and $y := (q_{n,k})_n$. Then, since if $|m| \leqslant |n|$ then

$$
\begin{aligned}
|p_{n,k} - p_{m,k}| &= 2^{-|n|} \cdot \left|\theta(\overline{|n|}, k) - \theta(\overline{|m|}, k) \cdot 2^{|n| \dot{-} |m|}\right| \\
&= 2^{-|n|} \cdot \left|\theta(\overline{|n|}, k) - \lfloor \theta(\overline{|n|}, k)/2^{|n| \dot{-} |m|} \rfloor \cdot 2^{|n| \dot{-} |m|}\right| \\
&< 2^{-|n|} \cdot 2^{|n| \dot{-} |m|} = 2^{-|m|} < 2^{-|n|} + 2^{-|m|}
\end{aligned}
$$

and

$$
\begin{aligned}
&|q_{n,k} - q_{m,k}| \\
&\quad= 2^{-|n|} \cdot \left|\theta(\overline{|n|}, k) + 1 - (\theta(\overline{|m|}, k) + 1) \cdot 2^{|n| \dot{-} |m|}\right| \\
&\quad\leqslant 2^{-|n|} \cdot \left|\theta(\overline{|n|}, k) - \theta(\overline{|m|}, k) \cdot 2^{|n| \dot{-} |m|}\right| + 2^{-|n|} \cdot \left|1 - 2^{|n| \dot{-} |m|}\right| \\
&\quad< 2^{-|n|} \cdot \left|\theta(\overline{|n|}, k) - \lfloor \theta(\overline{|n|}, k)/2^{|n| \dot{-} |m|} \rfloor \cdot 2^{|n| \dot{-} |m|}\right| + 2^{-|n|} \\
&\quad< 2^{-|n|} \cdot 2^{|n| \dot{-} |m|} + 2^{-|n|} = 2^{-|n|} + 2^{-|m|},
\end{aligned}
$$

we have $x, y \in \mathbf{R}(\mathcal{U})$. Moreover, since $|p_{n,k} - q_{n,k}| \leqslant 2^{-|n|} < 2^{-|n|+1}$, we have $x = y$. Let α be the function defined as in Proposition 15. Then, since $\hat{p}_{n,k} = p_{n,k}$ and $\hat{q}_{n,k} = q_{n,k}$, we have

$$
f(x) \leqslant \varphi\left(p_{\alpha(\overline{|k|+1}),k}, \overline{|k| + 2}\right) + 2^{-|k|-1} \leqslant 2^{-|k|-1} + 2^{-|k|-1} = 2^{-|k|},
$$

and

$$
f(x) = f(y) \geqslant \varphi\left(q_{\alpha(\overline{|k|+1}),k}, \overline{|k| + 2}\right) - 2^{-|k|-1} \leqslant -2^{-|k|}. \qquad \square
$$

References

1. Errett Bishop, *Foundations of Constructive Mathematics*, McGraw-Hill, New York, 1967.
2. Errett Bishop and Douglas Bridges, *Constructive Analysis*, Springer, Berlin, 1985.
3. Peter Clote, *Computational models and functional algebras*, In E.R. Griffor ed., Handbook of Computability Theory, North-Holland, Amsterdam, 1999.
4. Alan Cobham, *The intrinsic computational difficulty of functions*, In Y. Bar-Hillel ed., Logic, Methodology and Philosophy of Science II, North-Holland, Amsterdam, 24–30, 1965.
5. P. Csillag, *Eine Bemerkung zur Auflosung der eingeschachtelten Rekursion*, Acta Sci. Math. Szeged. 11, 169–173, 1947.
6. Hajime Ishihara, *Function algebraic characterizations of the polytime functions*, Comput. Complexity 8, 346–356, 1999.
7. Hajime Ishihara, *Feasibly constructive analysis*, Sūrikaisekikenkyūsho Kōkyūroku 1169, 76–83, 2000.
8. Hajime Ishihara, *Constructive reverse mathematics: compactness properties*, In L. Crosilla and P. Schuster eds., From Sets and Types to Analysis and Topology: Towards Practicable Foundations for Constructive Mathematics, Oxford University Press, Oxford, 245–267, 2005.

9. Neil D. Jones, *Computability and Complexity: From a Programming Perspective*, MIT Press, Cambridge, 1997.
10. László Kalmár, *Egyszerü példa eldönthetetlen aritmetikai problémára*, Mate és Fizikai Lapok. 50, 1–23, 1943.
11. Ker-I Ko, *Complexity Theory of Real Functions*, Birkhäuser, Boston, 1991.
12. Lars Kristiansen, *Neat function algebraic characterizations of* LOGSPACE *and* LINSPACE, Comput. Complexity 14, 72–88, 2005.
13. Marian B. Pour-El and Jonathan I. Richards, *Computability in Analysis and Physics*, Springer, New York, Berlin, 1989.
14. Harvey E. Rose, *Subrecursion: Functions and Hierarchies*, Oxford University Press, Oxford, 1984.
15. Anne S. Troelstra and Dirk van Dalen, *Constructivism in Mathematics, Vol. I. An Introduction*, North-Holland, Amsterdam, 1988.
16. Klaus Weihrauch, *Computable Analysis*, Springer, Berlin, 2000.

100 Years of Zermelo's Axiom of Choice: What was the Problem with It?

Per Martin-Löf

Cantor conceived set theory in a sequence of six papers published in the Mathematische Annalen during the five year period 1879–1884. In the fifth of these papers, published in 1883 [1], he stated as a law of thought (Denkgesetz) that every set can be well-ordered or, more precisely, that it is always possible to bring any well-defined set into the form of a well-ordered set. Now to call it a law of thought was implicitly to claim self-evidence for it, but he must have given up that claim at some point, because in the 1890s he made an unsuccessful attempt at demonstrating the well-ordering principle [2].

The first to succeed in doing so was Zermelo [3], although, as a prerequisite of the demonstration, he had to introduce a new principle, which came to be called the principle of choice (Prinzip der Auswahl) respectively the axiom of choice (Axiom der Auswahl) in his two papers from 1908 [4, 5]. His first paper on the subject, published in 1904, consists of merely three pages, excerpted by Hilbert from a letter which he had received from Zermelo. The letter is dated 24 September 1904, and the excerpt begins by saying that the demonstration came out of discussions with Erhard Schmidt during the preceding week, which means that we can safely date the appearance of the axiom of choice and the demonstration of the well-ordering theorem to September 1904.

Brief as it was, Zermelo's paper gave rise to what is presumably the most lively discussion among mathematicians on the validity, or acceptability, of a mathematical axiom that has ever taken place. Within a couple of years, written contributions to this discussion had been published by Felix Bernstein, Schoenflies, Hamel, Hessenberg and Hausdorff in Germany, Baire, Borel, Hadamard, Lebesgue, Richard and Poincaré in France, Hobson, Hardy, Jourdain and Russell in England, Julius König in Hungary, Peano in Italy and Brouwer in the Netherlands [2, pp. 92–137]. Zermelo responded to those of these contributions that were critical, which was a majority, in a second paper from 1908. This second paper also contains a new proof of the well-ordering theorem, less intuitive or less perspicuous, it has to be

P. Martin-Löf (✉)
Department of Mathematics, University of Stockholm, Sweden
e-mail: pml@math.su.se

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_10, © Per Martin-Löf 2006. Previously published in *The Computer Journal* 49:3, pp. 345–350, Oxford University Press.

admitted, than the original proof, as well as a new formulation of the axiom of choice, a formulation which will play a crucial role in the following considerations.

Despite the strength of the initial opposition against it, Zermelo's axiom of choice gradually came to be accepted mainly because it was needed at an early stage in the development of several branches of mathematics, not only set theory, but also topology, algebra and functional analysis, for example. Towards the end of the thirties, it had become firmly established and was made part of the standard mathematical curriculum in the form of Zorn's lemma [6].

The intuitionists, on the other hand, rejected the axiom of choice from the very beginning. Baire, Borel and Lebesgue were all critical of it in their contributions to the correspondence which was published under the title Cinq lettres sur la théorie des ensembles in 1905 [7]. Brouwer's thesis from 1907 contains a section on the well-ordering principle in which it is treated in a dismissive fashion ("of course there is no motivation for this at all") and in which, following Borel [8], he belittles Zermelo's proof of it from the axiom of choice [9]. No further discussion of the axiom of choice seems to be found in either Brouwer's or Heyting's writings. Presumably, it was regarded by them as a prime example of a nonconstructive principle.

It therefore came as a surprise when, as late as in 1967, Bishop stated,

> A choice function exists in constructive mathematics, because a choice is *implied by the very meaning of existence* [10],

although, in the terminology that he himself introduced in the subsequent chapter, he ought to have said "choice operation" rather than "choice function". What he had in mind was clearly that the truth of

$$(\forall i : I)(\exists x : S)A(i, x) \to (\exists f : I \to S)(\forall i : I)A(i, f(i))$$

and even, more generally,

$$(\forall i : I)(\exists x : S_i)A(i, x) \to (\exists f : \prod_{i:I} S_i)(\forall i : I)A(i, f(i))$$

becomes evident almost immediately upon remembering the Brouwer-Heyting-Kolmogorov interpretation of the logical constants, which means that it might as well have been observed already in the early thirties. And it is this intuitive justification that was turned into a formal proof in constructive type theory, a proof that effectively uses the strong rule of ∃-elimination that it became possible to formulate as a result of having made the proof objects appear in the system itself and not only in its interpretation.

In 1975, soon after Bishop's vindication of the constructive axiom of choice, Diaconescu proved that, in topos theory, the law of excluded middle follows from the axiom of choice [11]. Now, topos theory being an intuitionistic theory, albeit impredicative, this is on the surface of it incompatible with Bishop's observation because of the constructive inacceptability of the law of excluded middle. There is therefore a need to investigate how the constructive axiom of choice, validated by the Brouwer-Heyting-Kolmogorov interpretation, is related to Zermelo's axiom of choice on the one hand and to the topos-theoretic axiom of choice on the other.

To this end, using constructive type theory as our instrument of analysis, let us simply try to prove Zermelo's axiom of choice. This attempt should of course fail, but in the process of making it we might at least be able to discover what it is that is really objectionable about it. So what was Zermelo's axiom of choice? In the original paper from 1904, he gave to it the following formulation,

> *Jeder Teilmenge M' denke man sich ein beliebiges Element m'_1 zugeordnet, das in M' selbst vorkommt und das "ausgezeichnete" Element von M' genannt werden möge* [3, p. 514].

Here M' is an arbitrary subset, which contains at least one element, of a given set M. What is surprising about this formulation is that there is nothing objectionable about it from a constructive point of view. Indeed, the distinguished element m'_1 can be taken to be the left projection of the proof of the existential proposition $(\exists x : M)M'(x)$, which says that the subset M' of M contains at least one element. This means that one would have to go into the demonstration of the well-ordering theorem in order to determine exactly what are its nonconstructive ingredients.

Instead of doing that, I shall turn to the formulation of the axiom of choice that Zermelo favoured in his second paper on the well-ordering theorem from 1908,

> Axiom. *Eine Menge S, welche in eine Menge getrennter Teile A, B, C, ... zerfällt, deren jeder mindestens ein Element enthält, besitzt mindestens eine Untermenge S_1, welche mit jedem der betrachteten Teile A, B, C, ... genau ein Element gemein hat* [4, p. 110].

Formulated in this way, Zermelo's axiom of choice turns out to coincide with the multiplicative axiom, which Whitehead and Russell had found indispensable for the development of the theory of cardinals [12, 13]. The type-theoretic rendering of this formulation of the axiom of choice is straightforward, once one remembers that a basic set in the sense of Cantorian set theory corresponds to an extensional set, that is, a set equipped with an equivalence relation, in type theory, and that a subset of an extensional set is interpreted as a propositional function which is extensional with respect to the equivalence relation in question. Thus the data of Zermelo's 1908 formulation of the axiom of choice are a set S, which comes equipped with an equivalence relation $=_S$, and a family $(A_i)_{i:I}$ of propositional functions on S satisfying the following properties,

(1) $x =_S y \to (A_i(x) \leftrightarrow A_i(y))$ (extensionality),
(2) $i =_I j \to (\forall x : S)(A_i(x) \leftrightarrow A_j(x))$ (extensionality of the dependence on the index),
(3) $(\exists x : S)(A_i(x)\ \&\ A_j(x)) \to i =_I j$ (mutual exclusiveness),
(4) $(\forall x : S)(\exists i : I)A_i(x)$ (exhaustiveness),
(5) $(\forall i : I)(\exists x : S)A_i(x)$ (nonemptiness).

Given these data, the axiom guarantees the existence of a propositional function S_1 on S such that

(6) $x =_S y \to (S_1(x) \leftrightarrow S_1(y))$ (extensionality),
(7) $(\forall i : I)(\exists! x : S)(A_i \cap S_1)(x)$ (uniqueness of choice).

The obvious way of trying to prove (6) and (7) from (1)–(5) is to apply the type-theoretic (constructive, intensional) axiom of choice to (5), so as to get a function $f : I \to S$ such that

$$(\forall i : I)A_i(f(i)),$$

and then define S_1 by the equation

$$S_1 = \{f(j) \mid j : I\} = \{x \mid (\exists j : I)(f(j) =_S x)\}.$$

Defined in this way, S_1 is clearly extensional, which is to say that it satisfies (6). What about (7)? Since the proposition

$$(A_i \cap S_1)(f(i)) = A_i(f(i)) \;\&\; S_1(f(i))$$

is clearly true, so is

$$(\forall i : I)(\exists x : S)(A_i \cap S_1)(x),$$

which means that only the uniqueness condition remains to be proved. To this end, assume that the proposition

$$(A_i \cap S_1)(x) = A_i(x) \;\&\; S_1(x)$$

is true, that is, that the two propositions

$$\begin{cases} A_i(x), \\ S_1(x) = (\exists j : I)(f(j) =_S x) \end{cases}$$

are both true. Let $j : I$ satisfy $f(j) =_S x$. Then, since $(\forall i : I)A_i(f(i))$ is true, so is $A_j(f(j))$. Hence, by the extensionality of A_j with respect to $=_S$, $A_j(x)$ is true, which, together with the assumed truth of $A_i(x)$, yields $i =_I j$ by the mutual exclusiveness of the family of subsets $(A_i)_{i:I}$. At this stage, in order to conclude that $f(i) =_S x$, we need to know that the choice function f is extensional, that is, that

$$i =_I j \to f(i) =_S f(j).$$

This, however, is not guaranteed by the constructive, or intensional, axiom of choice which follows from the strong rule of $\exists$-elimination in type theory. Thus our attempt to prove Zermelo's axiom of choice has failed, as was to be expected.

On the other hand, we have succeeded in proving that Zermelo's axiom of choice follows from the extensional axiom of choice

$$(\forall i : I)(\exists x : S)A_i(x) \to (\exists f : I \to S)(\mathrm{Ext}(f) \;\&\; (\forall i : I)A_i(f(i))),$$

which I shall call ExtAC, where

$$\text{Ext}(f) = (\forall i, j : I)(i =_I j \rightarrow f(i) =_S f(j)).$$

The only trouble with it is that it lacks the evidence of the intensional axiom of choice, which does not prevent one from investigating its consequences, of course.

Theorem 1 *The following are equivalent in constructive type theory:*

(i) *The extensional axiom of choice.*

(ii) *Zermelo's axiom of choice.*

(iii) *Epimorphisms split, that is, every surjective extensional function has an extensional right inverse.*

(iv) *Unique representatives can be picked from the equivalence classes of any given equivalence relation.*

Of these four equivalent statements, (iii) is the topos-theoretic axiom of choice, which is thus equivalent, not to the constructively valid type-theoretic axiom of choice, but to Zermelo's axiom of choice.

Proof We shall prove the implications (i)→(ii)→(iii)→(iv)→(i) in this order.

(i)→(ii). This is precisely the result of the considerations prior to the formulation of the theorem.

(ii)→(iii). Let $S, =_S$ and $I, =_I$ be two extensional sets, and let $f : S \rightarrow I$ be an extensional and surjective mapping between them. By definition, put

$$A_i = f^{-1}(i) = \{x | f(x) =_I i\}.$$

Then

(1) $x =_S y \rightarrow (A_i(x) \leftrightarrow A_i(y))$

by the assumed extensionality of f,

(2) $i =_I j \rightarrow (\forall x : S)(A_i(x) \leftrightarrow A_j(x))$

since $f(x) =_I i$ is equivalent to $f(x) =_I j$ provided that $i =_I j$,

(3) $(\exists x : S)(A_i(x) \;\&\; A_j(x)) \rightarrow i =_I j$

since $f(x) =_I i$ and $f(x) =_I j$ together imply $i =_I j$,

(4) $(\forall x : S)(\exists i : I)A_i(x)$

since $A_{f(x)}(x)$ for any $x : S$, and

(5) $(\forall i : I)(\exists x : S)A_i(x)$

by the assumed surjectivity of the function f. Therefore we can apply Zermelo's axiom of choice to get a subset S_1 of S such that

$$(\forall i : I)(\exists! x : S)(A_i \cap S_1)(x).$$

The constructive, or intensional, axiom of choice, to which we have access in type theory, then yields $g : I \to S$ such that $(A_i \cap S_1)(g(i))$, that is,

$$(f(g(i)) =_I i) \;\&\; S_1(g(i)),$$

so that g is a right inverse of f, and

$$(A_i \cap S_1)(x) \to g(i) =_S x.$$

It remains only to show that g is extensional. So assume $i, j : I$. Then we have

$$(A_i \cap S_1)(g(i))$$

as well as

$$(A_j \cap S_1)(g(j))$$

so that, if also $i =_I j$,

$$(A_i \cap S_1)(g(j))$$

by the extensional dependence of A_i on the index i. The uniqueness property of $A_i \cap S_1$ permits us to now conclude $g(i) =_S g(j)$ as desired.

(iii)→(iv). Let I be a set equipped with an equivalence relation $=_I$. Then the identity function on I is an extensional surjection from I, Id_I to $I, =_I$, since any function is extensional with respect to the identity relation. Assuming that epimorphisms split, we can conclude that there exists a function $g : I \to I$ such that

$$g(i) =_I i$$

and

$$i =_I j \to \mathrm{Id}_I(g(i), g(j)),$$

which is to say that g has the miraculous property of picking a unique representative from each equivalence class of the given equivalence relation $=_I$.

(iv)→(i). Let $I, =_I$ and $S, =_S$ be two sets, each equipped with an equivalence relation, and let $(A_i)_{i:I}$ be a family of extensional subsets of S,

$$x =_S y \to (A_i(x) \leftrightarrow A_i(y)),$$

which depends extensionally on the index i,

$$i =_I j \to (\forall x : S)(A_i(x) \leftrightarrow A_j(x)).$$

Furthermore, assume that

$$(\forall i : I)(\exists x : S)A_i(x)$$

holds. By the intensional axiom of choice, valid in constructive type theory, we can conclude that there exists a choice function $f : I \to S$ such that

$$(\forall i : I)A_i(f(i)).$$

This choice function need not be extensional, of course, unless $=_I$ is the identity relation on the index set I. But, applying the miraculous principle of picking a unique representative of each equivalence class to the equivalence relation $=_I$, we get a function $g : I \to I$ such that

$$g(i) =_I i$$

and

$$i =_I j \to \mathrm{Id}_I(g(i), g(j)).$$

Then $f \circ g : I \to S$ becomes extensional,

$$i =_I j \to \mathrm{Id}_I(g(i), g(j)) \to \underbrace{f(g(i))}_{(f\circ g)(i)} =_S \underbrace{f(g(j))}_{(f\circ g)(j)}.$$

Moreover, from $(\forall i : I)A_i(f(i))$, it follows that

$$(\forall i : I)A_{g(i)}(f(g(i))).$$

But

$$g(i) =_I i \to (\forall x : S)(A_{g(i)}(x) \leftrightarrow A_i(x)),$$

so that

$$(\forall i : I)A_i(\underbrace{f(g(i))}_{(f\circ g)(i)}).$$

Hence $f \circ g$ has become an extensional choice function, which means that the extensional axiom of choice is satisfied.

Another indication that the extensional axiom of choice is the correct type-theoretic rendering of Zermelo's axiom of choice comes from constructive set theory. Peter Aczel has shown how to interpret the language of Zermelo-Fraenkel set theory in constructive type theory, this interpretation being the natural constructive version of the cumulative hierarchy, and investigated what set-theoretical principles become validated under that interpretation [14]. But one may also ask, conversely, what principle, or principles, have to be adjoined to constructive type theory in order to validate a specific set-theoretical axiom. In particular, this may be asked about the formalized version of the axiom of choice that Zermelo made part of his own axiomatization of set theory. The answer is as follows.

Theorem 2 *When constructive type theory is strengthened by the extensional axiom of choice, the set-theoretical axiom of choice becomes validated under the Aczel interpretation.*

Proof The set-theoretical axiom of choice says that, for any two iterative sets α and β and any relation R between iterative sets,

$$(\forall x \in \alpha)(\exists y \in \beta)R(x, y) \to (\exists \phi : \alpha \to \beta)(\forall x \in \alpha)R(x, \phi(x)).$$

The Aczel interpretation of the left-hand member of this implication is

$$(\forall x : \bar{\alpha})(\exists y : \bar{\beta})R(\tilde{\alpha}(x), \tilde{\beta}(y)),$$

which yields

$$(\exists f : \bar{\alpha} \to \bar{\beta})(\forall x : \bar{\alpha})R(\tilde{\alpha}(x), \tilde{\beta}(f(x)))$$

by the type-theoretic axiom of choice. Now, put

$$\phi = \{\langle \tilde{\alpha}(x), \tilde{\beta}(f(x))\rangle | x : \bar{\alpha}\}$$

by definition. We need to prove that ϕ is a function from α to β in the sense of constructive set theory, that is,

$$\tilde{\alpha}(x) = \tilde{\alpha}(x') \to \tilde{\beta}(f(x)) = \tilde{\beta}(f(x')).$$

Define the equivalence relations

$$(x =_{\bar{\alpha}} x') = (\tilde{\alpha}(x) = \tilde{\alpha}(x'))$$

and

$$(y =_{\bar{\beta}} y') = (\tilde{\beta}(y) = \tilde{\beta}(y'))$$

on $\bar{\alpha}$ and $\bar{\beta}$, respectively. By the extensional axiom of choice in type theory, the choice function $f : \bar{\alpha} \to \bar{\beta}$ can be taken to be extensional with respect to these two equivalence relations,

$$x =_{\bar{\alpha}} x' \rightarrow f(x) =_{\bar{\beta}} f(x'),$$

which ensures that ϕ, defined as above, is a function from α to β in the sense of constructive set theory.

Corollary *When constructive type theory (including one universe and the well-ordering operation) is strengthened by the extensional axiom of choice, it interprets all of* ZFC.

Proof We already know from Aczel that ZF is equivalent to CZF + EM [14, p. 59]. Hence ZFC is equivalent to CZF + EM + AC, where AC is the set-theoretical axiom of choice. But the Aczel translation interprets CZF in CTT and, by the previous theorem, AC in CTT + ExtAC. Therefore ZFC = CZF + EM + AC is interpretable in CTT + EM + ExtAC. It now only remains to appeal to the type-theoretical version of Diaconescu's theorem,[1] according to which the law of excluded middle follows from the extensional axiom of choice in the context of constructive type theory, in order to arrive at the final conclusion that ZFC is interpretable in CTT + ExtAC.[2]

When Zermelo's axiom of choice is formulated in the context of constructive type theory instead of Zermelo-Fraenkel set theory, it appears as ExtAC, the extensional axiom of choice

$$(\forall i : I)(\exists x : S)A(i, x) \rightarrow (\exists f : I \rightarrow S)(\mathrm{Ext}(f) \ \&\ (\forall i : I)A(i, f(i))),$$

where

$$\mathrm{Ext}(f) = (\forall i, j : I)(i =_I j \rightarrow f(i) =_S f(j)),$$

and it then becomes manifest what is the problem with it: it breaks the principle that you cannot get something from nothing. Even if the relation $A(i, x)$ is extensional with respect to its two arguments, the truth of the antecedent $(\forall i : I)(\exists x : S)A(i, x)$, which does guarantee the existence of a choice function $f : I \rightarrow S$ satisfying $(\forall i : I)A(i, f(i))$, is not enough to guarantee the extensionality of the choice function, that is, the truth of $\mathrm{Ext}(f)$. Thus the problem with Zermelo's axiom of choice is not the existence of the choice function but its extensionality, and this is not visible within an extensional framework, like Zermelo-Fraenkel set theory, where all functions are by definition extensional.

If we want to ensure the extensionality of the choice function, the antecedent clause of the extensional axiom of choice has to be strengthened. The natural way of doing this is to replace ExtAC by AC!, the axiom of unique choice, or no choice,

[1] S. Lacas and B. Werner, Which choices imply the Excluded Middle? About Diaconescu's trick in Type Theory, Unpublished, 1999, pp. 9–10. I am indebted to Jesper Carlström for providing me with this reference.

[2] In an earlier version of this paper, I gave two alternative proofs of the previous corollary. However, as has been kindly pointed out to me by Peter Schuster, the first of these contained an error, whence only what was originally the second proof now remains.

$$(\forall i : I)(\exists! x : S)A(i, x) \to (\exists f : I \to S)(\mathrm{Ext}(f)\ \&\ (\forall i : I)A(i, f(i))),$$

which is as valid as the intensional axiom of choice. Indeed, assume $(\forall i : I)(\exists! x : S)A(i, x)$ to be true. Then, by the intensional axiom of choice, there exists a choice function $f : I \to S$ satisfying $(\forall i : I)A(i, f(i))$. Because of the uniqueness condition, such a function $f : I \to S$ is necessarily extensional. For suppose that $i, j : I$ are such that $i =_I j$ is true. Then $A(i, f(i))$ and $A(j, f(j))$ are both true. Hence, by the extensionality of $A(i, x)$ in its first argument, so is $A(i, f(j))$. The uniqueness condition now guarantees that $f(i) =_S f(j)$, that is, that $f : I \to S$ is extensional. The axiom of unique choice AC! may be considered as the valid form of extensional choice, as opposed to ExtAC, which lacks justification.

The inability to distinguish between the intensional and the extensional axiom of choice has led to one's taking the need for the axiom of choice in proving that the union of a countable sequence of countable sets is again countable, that the real numbers, defined as Cauchy sequences of rational numbers, are Cauchy complete, etc., as a justification of Zermelo's axiom of choice. As Zermelo himself wrote towards the end of the short paper in which he originally introduced the axiom of choice,

> Dieses logische Prinzip lät sich zwar nicht auf ein noch einfacheres zurückführen, wird aber in der mathematischen Deduktion überall unbedenklich angewendet [3, p. 516].

What Zermelo wrote here about the omnipresent, and often subconscious, use of the axiom of choice in mathematical proofs is incontrovertible, but it concerns the constructive, or intensional, version of it, which follows almost immediately from the strong rule of existential elimination. It cannot be taken as a justification of his own version of the axiom of choice, including as it does the extensionality of the choice function.

Within an extensional foundational framework, like topos theory or constructive set theory, it is not wholly impossible to formulate a counterpart of the constructive axiom of choice, despite of its intensional character, but it becomes complicated. In topos theory, there is the axiom that there are enough projectives, which is to say that every object is the surjective image of a projective object, and, in constructive set theory, Aczel has introduced the analogous axiom that every set has a base [15]. Roughly speaking, this is to say that every set is the surjective image of a set for which the axiom of choice holds. The technical complication of these axioms speaks to my mind for an intensional foundational framework, like constructive type theory, in which the intuitive argument why the axiom of choice holds on the Brouwer-Heyting-Kolmogorov interpretation is readily formalized, and in which whatever extensional notions that are needed can be built up, in agreement with the title of Martin Hofmann's thesis, Extensional Constructs in Intensional Type Theory [16]. Extensionality does not come for free.

Finally, since this is only a couple of weeks from the centenary of Zermelo's first formulation of the axiom of choice, it may not be out of place to remember the crucial role it has played for the formalization of both Zermelo-Fraenkel set theory and constructive type theory. In the case of set theory, there was the need for

Zermelo of putting his proof of the well-ordering theorem on a formally rigorous basis, whereas, in the case of type theory, there was the intuitively convincing argument which made the axiom of choice evident on the constructive interpretation of the logical operations, an argument which nevertheless could not be faithfully formalized in any then existing formal system.

References

1. G. Cantor, Über unendliche lineare Punktmannigfaltigkeiten. Nr. 5, *Math. Annalen*, Vol. 21, 1883, pp. 545-591. Reprinted in *Gesammelte Abhandlungen*, Edited by E. Zermelo, Springer-Verlag, Berlin, 1932, pp. 165–208.
2. G. H. Moore, *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, Springer-Verlag, New York, 1982, p. 51.
3. E. Zermelo, Beweis, da jede Menge wohlgeordnet werden kann. (Aus einem an Herrn Hilbert gerichteten Briefe.), *Math. Annalen*, Vol. 59, 1904, pp. 514–516.
4. E. Zermelo, Neuer Beweis für die Möglichkeit einer Wohlordnung, *Math. Annalen*, Vol. 65, 1908, pp. 107–128.
5. E. Zermelo, Untersuchungen über die Grundlagen der Mengenlehre. I, *Math. Annalen*, Vol. 65, 1908, pp. 261–281.
6. M. Zorn, A remark on method in transfinite algebra, *Bull. Amer. Math. Soc.*, Vol. 41, 1935, pp. 667–670.
7. R. Baire, É. Borel, J. Hadamard and H. Lebesgue, Cinq lettres sur la théorie des ensembles, *Bull. Soc. Math. France*, Vol. 33, 1905, pp. 261–273.
8. É. Borel, Quelques remarques sur les principes de la théorie des ensembles, *Math. Annalen*, Vol. 60, 1905, pp. 194–195.
9. L. E. J. Brouwer, *Over de grondslagen der wiskunde*, Maas & van Suchtelen, Amsterdam, 1907. English translation in *Collected Works*, Vol. 1, Edited by A. Heyting, North-Holland, Amsterdam, 1975, pp. 11–101.
10. E. Bishop, *Foundations of Constructive Analysis*, McGraw-Hill, New York, 1967, p. 9.
11. R. Diaconescu, Axiom of choice and complementation, *Proc. Amer. Math. Soc.*, Vol. 51, 1975, pp. 176–178.
12. A. N. Whitehead, On cardinal numbers, *Amer. J. Math.*, Vol. 24, 1902, pp. 367–394.
13. B. Russell, On some difficulties in the theory of transfinite numbers and order types, *Proc. London Math. Soc.*, Ser. 2, Vol. 4, 1906, pp. 29–53.
14. P. Aczel, The type theoretic interpretation of constructive set theory, *Logic Colloquium '77*, Edited by A. Macintyre, L. Pacholski and J. Paris, North-Holland, Amsterdam, 1978, pp. 55–66.
15. P. Aczel, The type theoretic interpretation of constructive set theory: choice principles, *The L. E. J. Brouwer Centenary Symposium*, Edited by A. S. Troelstra and D. van Dalen, North-Holland, Amsterdam, 1982, pp. 1–40.
16. M. Hofmann, *Extensional Constructs in Intensional Type Theory*, Springer-Verlag, London, 1997.

Intuitionism and the Anti-Justification of Bivalence

Peter Pagin

Abstract Dag Prawitz has argued [12] that it is possible intuitionistically to prove the validity of '$A \rightarrow$ there is a proof of $\ulcorner A \urcorner$' by induction over formula complexity, provided we observe an object language/meta-language distinction. In the present paper I mainly argue that if the object language with its axioms and rules can be represented as a formal system, then the proof fails. I also argue that if this restriction is lifted, at each level of the language hierarchy, then the proof can go through, but at the expense of virtually reducing the concept of a proof to that of truth in a non-constructive sense.

1 Background

A couple of years ago [9] I argued against Dag Prawitz and Michael Dummett that the principle of bivalence, i.e. the principle that every sentence is either true or false, cannot have both a metaphysical and a meaning-theoretical significance. This *double significance view* is a conjunction of two claims. The first claim is that acceptance of bivalence for a particular area of discourse is the principal mark of a *realist* view of that area. Correspondingly, rejection of bivalence is the principal mark of an *anti-realist* view. Even though there are other ingredients in realism and anti-realism, and even though there are shades of realism and anti-realism, the stand on bivalence is the main dividing line.[1]

The bivalence criterion was not meant to introduce a new idea distinct from traditional metaphysics, but as a way of articulating traditional metaphysical ideas in a clearer and more precise way.[2] The realist idea to be captured by the bivalence

P. Pagin (✉)
Department of Philosophy, Stockheelm University, Stockheelm, Sweden
e-mail: peter.pagin@philosophy.su.se

A predecessor of the paper was presented at the Uppsala conference in August 2004. I benefited there especially from comments by John Burgess, Dag Prawitz, Stewart Shapiro and Göran Sundholm. In addition, the paper improved significantly thanks to comments from an anonymous referee. My overall philosophical debt to Dag Prawitz should be evident from the paper.

[1] For an overview and classification of positions, see [3].

[2] The general idea is spelled out in the last chapter of [4].

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341, DOI 10.1007/978-1-4020-8926-8_11, © Springer Science+Business Media B.V. 2009

criterion is that of a determinate and mind-independent reality. Mind-independence is here equal to independence of being *knowable* (by humans or at least by finite minds). A certain domain D of reality is *mind-independent* if it is possible that there be unknowable facts in D. D is also *determinate* if all states of affairs of D either obtain or don't obtain.

To a reality domain D, which may be a domain such as the past, mathematical reality, or counterfactual reality, there corresponds an area of discourse, A_D, of sentences describing states of affairs in D. Assuming that everything in D is expressible in A_D, we can say that a state of affairs in D obtains just in case a sentence s in A_D that describes it is *true*, and it fails to obtain if s is false. It will also fail to obtain if s is neither true nor false, but I shall not here take account of that option. Therefore, every state of affairs in D either obtains or does not obtain just in case every sentence in the corresponding discourse area A_D is either true or false. Hence, the following general equivalence of bivalent reality:

(BR) A domain D is determinate iff A_D is bivalent

The realism issue for a domain D will not arise if we know that for any sentence s in A_D it is either the case that we can know that s is true or the case that we can know that s is false. For, if we know that, then we do know that D is *not* mind-independent. An example might be, for each person, the domain of present sensations, reflected in sentences like 'I am in pain'. Disputed domains are those where we lack any such guarantee. For such disputed domains we have two alternatives. Either we can hold on to bivalence, in which case we will also hold that domain to be mind-independent, for then every sentence of that area of discourse is true or is false regardless of whether we can know it or not. This is realism. Or we can reject mind-independence for the domain, and as a consequence we will then have to reject bivalence, for if a sentence is true just in case knowably true, and false just in case it is knowably false, and there is no guarantee that it be either knowably true or knowably false, then there is no guarantee that it is either true or false. So the issue of determinacy is closely connected with that of mind dependence. Let's say that

(KA) An area of discourse A_D is *knowable* iff it holds of every sentence s in A_D that if s is true, then it can be known that s is true.

I shall here take falsity to be equivalent with truth of negation, so we don't need a separate clause about knowability of falsity. We can now define mind-independence:

(DK) A domain D is *mind-independent* iff A_D is not knowable

For disputed domains, determinacy implies mind-independence, and acceptance of bivalence for the corresponding area of discourse amounts to affirming both. We can sum up the first conjunct of the double significance view as

(B1) A disputed domain D is *metaphysically real* iff A_D is bivalent.

The second conjunct in the double significance view is the claim that there are *meaning-theoretical reasons* for rejecting bivalence. One can put it by saying that there is a meaning-theoretical *anti-justification* of bivalence. Rejecting bivalence here does not consist in claiming that there are sentences that are neither true nor false, but rather in claiming that there are reasons for believing that not every sentence is either true or false. By classical logic, if not every sentence is either true or false, then some sentence is neither. However, this does not hold by intuitionistic logic, and that is what counts in the present context, since only intuitionistic logic is justifiable according to the meaning-theoretical requirements.

We can then state the second conjunct of the double significance view as

(B2) For any disputed domain D, there are meaning-theoretical reasons for believing that A_D is not bivalent.

I turn now to these meaning-theoretical reasons. What I shall call *Dummett's argument* is an argument to the conclusion that meaning cannot consist in truth conditions, at least as truth conditions are normally understood. In a very condensed form, it can be set out as follows:[3]

1. Knowledge of the meaning of a sentence is publicly manifestable.
2. Public manifestation of knowledge of the meaning of a sentence consists in exercising the ability to tell whether the central semantic concept applies to that sentence or not.
3. If the central semantic concept is the concept of truth (and therefore knowledge of meaning is knowledge of truth conditions) and the sentence is not effectively decidable, then there is no ability to tell whether that concept applies to the sentence.
4. Hence knowledge of truth conditions is not (always) publicly manifestable.
5. Hence knowledge of meaning is not knowledge of truth conditions.
6. Hence meaning is not truth conditions.

This is the way I have interpreted Dummett. Granting the argument for the sake of discussion, we have the requirement that the central semantic concept must be *decidable*. I granted this much in [9], and I also granted that best alternative semantic concept would be the relation *x is a proof of y*.[4] I assumed that this relation is decidable. If it is decidable, a semantic theory that has it as the central semantic concept would meet Dummett's manifestability requirement.

Further, I assumed that the meanings of the logical constants are given in terms of what counts as a proof, or canonical proof, of sentences with those constants as main operator, as is done in the intuitionistic tradition. Then, if a sentence is logically valid just in case all sentences of the same logical form have a proof, instances of

[3] Extracted mainly from [1].
[4] In this case the concept is dyadic. What most directly compares with the monadic concept of *truth* is the relativized monadic concept *x is a proof of* ..., one for each purported proof x.

(EM) A or not A

are not logically valid.

Given the equivalence principle, or disquotation principle, i.e. the principle that every instance of

(EP) 'A' is true iff A

is true, or correct in some appropriate sense, and again equating falsity with truth of negation, it also holds that the principle of bivalence cannot be logically justified: it is not derivable from the meaning explanations of the logical constants together with the equivalence principle, because (EM) is not a logically valid schema.

As a conclusion, bivalence is not logically justifiable. This is the basis for the claim that bivalence is not a *valid* principle, or that there is no guarantee that it is valid, or again that there is no reason for believing in its validity. This amounts to a *weak* rejection of bivalence: we lack positive reasons for believing in its validity. The *strong* rejection is the claim that there are reasons for believing that it is *not* valid. Part of the reason for thinking that bivalence is not valid is precisely that there is no general justification of its validity. This is then a meaning-theoretical rejection, or meaning-theoretically motivated rejection, and this is the second conjunct of the double significance thesis.

Then putting (B1) and (B2) together, we have as a consequence of the double significance thesis that

(B3) For any disputed domain D, there are meaning-theoretical reasons to think that D is not metaphysically real.

(B3) is a remarkable claim. How can reflection over the requirements on language for communication give such a metaphysical result?

In [9] I argued that it cannot. There is a gap in the reasoning. The mere fact that bivalence cannot be *logically* justified, i.e. justified from the meanings of the logical constants, does not by itself imply that it isn't valid in the sense that every instance is *true*. Reality might be such as to make either A or *not* A true, for some or for every disputed domain, even if we don't have any general guarantee that this is the case, i.e. even if we don't have any proof that this holds. It might be that for some particular sentence s of an area A_D s is true but not provable/verifiable. Or it may be that for each s of A_D either s or its negation is in fact provable, even though this general fact itself is *not* provable.[5]

In order to close this gap, truth and provability need to be connected in a non-trivial way. The most natural candidate is something like

[5] Neil Tennant's view, in [13], is in agreement with this claim. According to Tennant, principles that are valid according to intuitionistic relevant logic are *analytic*, while others, like The Law of Excluded Middle, if valid at all, are not justifiable from meaning explanations and therefore *synthetic*.

(P) If A is true, then there is a proof of A

(where I write 'A' instead of 'the proposition that A'). In the present terminology, if (P) holds for a particular domain D, then D is mind-dependent, assuming that proofs are knowable. Thus, arguing for the general validity of (P), for any domain D, is in effect arguing that reality in general is mind-dependent.

On some definitions of truth, e.g.

(TP) A is true $=_{\text{def}}$ there is a proof of A

proposed by Prawitz as well as by Per Martin-Löf, (P) holds by definition.[6] But since we cannot simply *define* reality to be mind-dependent, we would need to start from a more neutral notion or property of truth. The equivalence principle itself is sufficient, if we decide to consider truth of sentences rather than propositions. Let's do that. Then, given some restrictions, e.g. because of reference failure or for the sake of avoiding semantic paradoxes, the equivalence principle should hold for any acceptable conception of truth. Accepting this, we would have, for any sentence $\ulcorner A \urcorner$,

(P*) If A, then there is a proof of $\ulcorner A \urcorner$

I argued in [9] that (P*) isn't intuitionistically acceptable.[7] First, it is not enough to give a semantics for some particular object language and show (P*) to hold for that object language (or a fragment of it), for what counts for the *realism* issue is whether bivalence and thus excluded middle holds for the *home* language, the language that is *used*. The home language may be bivalent and yet be the meta-language for a non-bivalent object language.[8] So (P*) has to be proved for the home language itself, or for some fragment of it that comprises the area of discourse under discussion.

Secondly, I argued that there cannot be an intuitionistically acceptable proof of the validity of (P*) for the home language. A proof of (P*) would, according to the intuitionistic meaning explanations of the logical constants, be a function F that, for any sentence s, as the antecedent of the instance of (P*), for a proof of s as argument, gives a proof of the consequent 'there is a proof of s' as value. Such a proof of the consequent would be a pair $\langle a, b \rangle$ such that a is an object and b a proof of 'x is a proof of s' given the assignment of a to x (alternatively, that a is a term and b a proof of 'a is a proof of s').[9]

---

[6] See [7, p. 11], [10, p. 8], [11, p. 85].

[7] In [9] (P) is called '(PRO)' and (P*) '(PRO*)'. I was also not making the distinction between proof of a sentence and proof of a proposition explicit.

[8] This can be achieved e.g. by the method of interpreting the intuitionistic language into a classical modal language, and is carried out in [9], Section 5.

[9] Hence, accepting (P*) simply as an *axiom*, in which case the one-step inference from zero premises is the (unreducible) proof, is not in accordance with the intuitionistic meaning explanations.

I argued that there cannot be such a general function, on two assumptions. The first assumption was Church's Thesis, according to which any decidable property is recursive, and the second assumption that the relation *x is a proof of y* is decidable. On these assumptions, since the language of arithmetic, together with the axioms of arithmetic, is part of the home language, Gödel's incompleteness results apply. The first incompleteness theorem implies that there is no exhaustive recursive relation *x is a proof of y*. Any recursive relation is representable in a formal system for arithmetic. Gödel showed that for any such formal system G we can construct a universally quantified sentence s of which there is no formal derivation in G, but which is informally proved by the observation that each instance of s is provable in G. Hence, no formula of a formal system can represent the relation *x is a proof of y*.

Hence, again, if the relation *x is a proof of y* is to be decidable, there is no well-defined collection of *all proofs*, and therefore no function F that could be a proof of (P*). Finally, applying (P) to (the universal closure of) (P*) itself, we get

$$\text{If (P*) is true, then there is a proof of (P*)} \tag{1}$$

By contraposition, from the result that there is no proof of (P*), we get the conclusion that (P*) is not true. Thus, from an intuitionistic perspective, (P*) should in fact be rejected.[10]

2 Professor Prawitz's Reply I: The Proof Relation

Professor Prawitz accepted the challenge. His reply, [12], has two main claims. First, the proof relation, *x is a proof of y*, cannot reasonably be regarded as decidable, and isn't required to be. Second, (P*) can in fact be proved. I shall look at the first claim in this section.

According to Prawitz [12, 306], the proof relation is not recursive. If the reason for this were only the incompleteness results, then there is an alternative view. You can think that there isn't one but many proof relations. A proof relation R is a subset of $P \times S$, the product of a domain P of proofs and a domain S of sentences, or of a product $P \times Q$ of a domain of proofs and a domain of propositions, depending on whether we think of proofs as related to sentences or to propositions. Let's assume it is sentences. Then we could say that for any proof relation R there is another proof relation $R' \subset P' \times S$, where $P \subset P'$. The restriction of R' to $P \times S$ is identical with R, but the expanded domain P' of proofs contains elements that by R' are proofs of sentences in the original domain S. We would say, as did Dummett in his reflections on Gödel [2], that the domain of proofs is *indefinitely extensible*. It would be open

[10] It has been objected, and was so objected at the Uppsala conference, that Church's Thesis is not intuitionistically valid. However, the reason for rejecting Church's Thesis comes from Brouwer's theory of real numbers as choice sequences (the theory of the creative subject). The problems for (P*) I point to arise already with arithmetic, where as far as I know intuitionists accept Church's Thesis.

now to claim that although each of the members in the hierarchy of proof relations *is* recursive, there is no largest relation, no comprehensive proof relation.

Prawitz seems to take the opposite view, that there *is* a comprehensive proof relation, which, consequently, is not recursive. His reasons are not only related to the incompleteness results, but directly to the requirements on proofs by the intuitionistic meaning explanations. A proof of an implication $A \to B$ is to be a function that transforms any proof of A into a proof of B, and it may clearly be difficult to determine of a particular construction that it has this property.[11]

However, if we give up the idea that the proof relation is decidable, what happens with the manifestation requirements that plays a crucial role in Dummett's meaning-theoretical argument? Prawitz says

> Similarly, knowing a method for transforming proofs to other proofs and knowing the meaning of an implication $A \to B$, I should be able to decide whether the method *as I know it* amounts to a direct verification of the implication. The method may in fact transform any proof of A into a proof of B, but one cannot in general require that I should know a procedure that allows me to establish this fact. However, knowing the meaning of an implication, I should be able to decide for a given method known to me whether *in virtue of its properties known to me* the method amounts to a direct verification of $A \to B$. The required capacity is something else than a capacity to decide for an arbitrary object whether it is a direct verification of a given sentence ([12, pp. 306–337], italics in the original).

Thus, according to Prawitz, since the proof relation isn't recursive, and hence not decidable in any clear sense, one cannot require that a speaker be able to manifest his understanding of a sentence by displaying an ability to decide for any given object, whether or not that object is a proof, or is a canonical proof, of the sentence. The requirement should be weaker.

What is the weaker requirement? Prawitz suggests that the speaker should be able to decide, for an object a and a sentence s whether a, *in virtue of the properties known to the speaker*, a is a canonical verification of s. But what does 'in virtue of the properties known to the speaker' mean here? On one interpretation the requirement is a strong as the original one. For I may be aware of the properties of the object that are in fact such that in virtue of having them, i.e. as a consequence of having those properties, the object is a canonical proof of a $A \to B$, even though I am not able to verify that this *is* a consequence of having them. I can see how the object is constructed out of certain operators, and in virtue of being so constructed it is a proof of $A \to B$, but this I am not at the time able to tell. Hence, on this interpretation it is not clear that Dummett's requirement is weakened at all, and if it is, certainly not sufficiently.

On another interpretation, the requirement is clearly too weak. On this interpretation the phrase 'in virtue of the properties known to the speaker' refers to *proof properties the speaker is aware of*. That is, if the speaker is aware that a has the property of transforming any proof of A into a proof of B, then the speaker should also

[11] Although I don't object to Prawitz's reason, it may be noted that it goes against an old intuitionistic view, expressed by Georg Kreisel by the aphorism 'We can recognize a proof when we see one' [6, 202].

know that a is a proof of $A \to B$. However, this is an almost empty requirement. If I have no idea of what it requires of an object to be a function that transforms any proof of A into a proof of B, then I will never be able to tell when something is a proof of $A \to B$, but if I am required only to say that something is a proof of $A \to B$ in case I am aware that it transforms any proof of A into a proof of B, then I do meet the manifestation requirements nonetheless.

So this interpretation is too weak. What is needed is some interpretation intermediate in strength between the first and the second, weak enough not to require impossible capacities, but strong enough to rule out truth conditional semantics. It is not clear that there is such an intermediate manifestability requirement to be had.

The rejection of the demand that the proof relation be decidable does, however, take away the basis of my 1998 argument against the provability of (P*). Is it enough to make it provable?

3 Professor Prawitz's Reply II: The Provability of (P*)

Professor Prawitz accepted my demand that (PRO*) be intuitionistically provable as holding of our home language. He says

> The upshot of this seems to be that the metaphysical relevance of the meaning-theoretical argument against bivalence for a certain language hinges upon two things: firstly, that the language is *our* language for speaking about the part of reality in question, and, secondly, that the meaning theory for this language adequately describes our use of this language, in particular what our sentences mean and when they are true ([12 309–310], italics in original).

Prawitz does hold on to the (TP) truth definition, by which truth is the same as the existence of proof. (P) is an immediate consequence of (TP), but Prawitz thinks that Tarski's condition of material adequacy, of which the equivalence principle (EP) is a special case, should be satisfied. Because of this, both (P) and (P*) should be correct for the home language.

On Prawitz's view, this demand can in fact be met, and he goes into some detail in indicating how this is to be done. First, Prawitz requires a strict separation of linguistic levels, in order to avoid the liar paradox, in the format

(S) There is no proof of (S).

Liar reasoning gives the result that there is and there is not a proof of (S). Therefore, we need an object language/meta-language separation, according to Prawitz. There are of course other ways to avoid the liar paradox, but in the present context the Tarskian is as good as any other, and I shall here follow Prawitz. He adopts the picture of the home language as made up of a hierarchy of object language and metalanguages. Three are needed for Prawitz's construction, an object language, a meta-language and a meta-meta-language, and I shall here simply call them *L*, *ML* and *MML*. As usual, *ML* is an extension of *L*, and *MML* an extension of *ML*.

I shall use a notation that is slightly different from Prawitz's. I shall use 'A', 'B' etc. as substitutional variables for sentences of *L*. I shall use '$P(\ldots, \ldots)$' as relational expression in *ML* denoting the proof relation between objects and sentences of *L*, and boldface '$\boldsymbol{P}$' as the corresponding expression of *MML* (denoting a relation between objects and sentences of *ML*). 'a', 'b' etc. are variables of *ML* ranging over a domain of proofs of sentences of *L*, and correspondingly '$\boldsymbol{a}$', '$\boldsymbol{b}$' etc. as variables in *MML* ranging over a domain of proofs of sentences of *ML*.

I shall use '$^\circ$' for forming structural-descriptive names in *ML* of expressions in *L*, and '$^\bullet$' for forming structural-descriptive names in *MML* of expressions in *ML*. Note that since *MML* is an extension of *ML*, *ML* names of expressions in *L* are also *MML* names of expressions in *L*. I shall use '$\forall$' and '$\exists$' as objectual quantifier symbols for *ML* and *MML* alike.

With this much machinery, we can state the formal counterpart to the (P*) schema as

$$A \to \exists a(P(a, A^\circ)) \tag{2}$$

which is a sentence schema of *ML*. The claim that each instance of (2) is provable belongs, as Prawitz points out, to *MML*. With the addition of '[...]' as a substitutional universal quantifier of *MML*, we can state that claim as

$$[A](\exists \boldsymbol{b})(\boldsymbol{P}(\boldsymbol{b}, (A \to \exists a(P(a, A^\circ)))^\bullet)) \tag{3}$$

which in the present notation corresponds to Prawitz's (PRO+):

> For any sentence A in English$*_L$, there is an α such that
> Proof(α, $A \to (\exists p)\Pi(p, \mu(A))$)).[12]

I suppose we could simplify things by having the substitutional quantifier already in *ML*. Then it would be sufficient to prove

$$[A](A \to \exists a(P(a, A^\circ))) \tag{4}$$

Be that as it may. According to Prawitz, proving (PRO+) is sufficient for answering the challenge. He gives a sketch of how (PRO+) can be proved by induction over complexity of *L* sentences, and carries out the induction step for conjunction. As he points out, for this to work we must add semantic clauses for the expression of the proof relation, i.e. '$P(\ldots, \ldots)$'. This must be done in accordance with the meaning explanations for logical constants of *L*. Thus, translated into the present notation, Prawitz says in the case of conjunction that a canonical proof of a sentence

[12] Here '$(\exists p)\Pi(p, \mu(A))$' is a name in the meta-meta-language English$*_{MML}$ of a sentence in the meta-language English$*_{ML}$ which says that there is a proof of the sentence A of the object language English$*_L$. '$A \to (\exists p)\Pi(p, \mu(A))$' is the structural-descriptive *MML* name of the corresponding *ML* conditional.

$$P(c, (A\&B)^\circ) \tag{5}$$

of *ML* is to be the same as a canonical proof of

$$P(\mathbb{L}(\text{RUN}\,(c)), A^\circ)\ \&\ P(\mathbb{R}(\text{RUN}\,(c)), B^\circ) \tag{6}$$

The reason is this. A canonical proof of a conjunction $(A\&B)^\circ$ in L (named in *ML*), is a pair $\langle a, b\rangle$, where a is a proof of A° and b is a proof of B°. A proof in general is a *method* for arriving at a canonical proof. To accommodate this idea we need a function which takes a proof, i.e. a method, as argument, and delivers the corresponding canonical proof as value. We also need to enrich *ML* with a function symbol denoting this function, and 'RUN$(\ldots)$' is such a function symbol. We also need symbols for the *projection functions* that extract the left and right members of a pair, respectively, and '$\mathbb{L}$' and '$\mathbb{R}$' are such function symbols (thus, '$\mathbb{L}(\langle a, b\rangle) = a$' is true in *ML*).

Then, equating (5) and (6) amounts to saying that a proof that something c is a proof of a conjunction is the *same* as a proof that the elements of the canonical counterpart of c are proofs of the conjuncts. The alternative would be to include as a separate inference rule for *MML* that we can infer (5) from (6), (and another that the converse holds as well). In that case, a canonical proof of a sentence of the form of (5) would proceed by way of one application of this inference rule from a canonical proof of a sentence of the form of (6). In this case, the difference is of small significance, but as we shall, see, it will be more important in the case of the quantifiers.

Given the equation of (5) and (6), Prawitz's idea of the induction step for conjunction is pretty straightforward. I shall sketch it here. For the details, the reader is referred to [12]. First, we assume we have a proof $\boldsymbol{h}$ in *MML* of a sentence $(A\&B)^\bullet$ in ML. Since $\boldsymbol{h}$ is a method for arriving at a canonical proof of $(A\&B)^\bullet$ and such a canonical proof is a pair, we can from $\boldsymbol{h}$ extract a proof $\boldsymbol{c}$ of $A^\bullet$ and a proof $\boldsymbol{d}$ of $B^\bullet$. Now we can make use of the induction hypothesis. We infer by the induction hypothesis that there is a proof $\boldsymbol{e}$ of $(A \to \exists a(P(a, A^\circ)))^\bullet$ and a proof $\boldsymbol{g}$ of $(B \to \exists b(P(b, B^\circ)))^\bullet$.

Since by definition a canonical proof of an implication is a function that takes any proof of the antecedent into a proof of the consequent, we can apply $\boldsymbol{e}$ to $\boldsymbol{c}$, and $\boldsymbol{e}(\boldsymbol{c})$ is then a proof of $(\exists a(P(a, A^\circ)))^\bullet$. Similarly, $\boldsymbol{g}(\boldsymbol{d})$ is a proof of $(\exists b(P(b, B^\circ)))^\bullet$.

By definition, a canonical proof of an existential sentence is a pair where the first element is a term that replaces the variable, and the second element is a proof of the embedded sentence after the substitution. Therefore, from $\boldsymbol{e}(\boldsymbol{c})$ we can extract a pair $\langle u, \boldsymbol{m}\rangle$, where $\boldsymbol{m}$ is a proof of $(P(u, A^\circ)^\bullet$. Similarly, from $\boldsymbol{g}(\boldsymbol{d})$ we can extract a pair $\langle v, \boldsymbol{n}\rangle$, where $\boldsymbol{n}$ is a proof of $(P(v, B^\circ)^\bullet$. By forming the pair $\langle \boldsymbol{m}, \boldsymbol{n}\rangle$ we get a proof of the conjunction $(P(u, A^\circ)\ \&\ P(v, B^\circ))^\bullet$.

Now we apply the equivalence of (5) and (6). In virtue of this equivalence, the proof $\langle \boldsymbol{m}, \boldsymbol{n}\rangle$ of $(P(u, A^\circ)\ \&\ P(v, B^\circ))^\bullet$ is also a proof of $(P(\langle u, v\rangle, (A\&B)^\circ))^\bullet$. Now we can form the pair of pairs $\langle\langle u, v\rangle, \langle \boldsymbol{m}, \boldsymbol{n}\rangle\rangle$, which is a proof of the existential sentence $(\exists a(P(a, (A\&B)^\circ)))^\bullet$.

So, given the induction hypothesis, from the assumption of having a proof $\boldsymbol{h}$ of $(A\&B)^{\bullet}$ in ML we have derived a proof of $(\exists a(P(a, (A\&B)^{\circ})))^{\bullet}$. Discharging the assumption with an implication introduction, we have a proof (depending on the induction hypothesis) of $(A\&B \to \exists a(P(a, (A\&B)^{\circ})))^{\bullet}$. The derivation we have carried out corresponds to a series of operations in *MML* on the assumed proof $\boldsymbol{h}$, building up a complex construction $\boldsymbol{q}(\boldsymbol{h})$, which is a term in *MML* denoting the proof of $(\exists a(P(a, (A\&B)^{\circ})))^{\bullet}$. Discharging the assumption corresponds to abstracting on the argument, and we therefore have $\lambda x\boldsymbol{q}(x)$ as a proof term of *MML* denoting the proof of $(A\&B \to \exists a(P(a, (A\&B)^{\circ})))^{\bullet}$. With the reasoning carried out in *MML* there is a proof expressible in *MML* of the *MML* sentence

$$\boldsymbol{P}(\lambda x\boldsymbol{q}(x), (A\&B \to \exists a(P(a, (A\&B)^{\circ})))^{\bullet}) \tag{7}$$

Existential generalization gives

$$\exists\boldsymbol{b}(\boldsymbol{P}(\boldsymbol{b}, (A\&B \to \exists a(P(a, (A\&B)^{\circ})))^{\bullet})) \tag{8}$$

which concludes the induction step, as carried out by Prawitz.

4 Problems with Prawitz's Construction

Is Prawitz's induction step for conjunction correct as it stands, and can it be extended to a complete induction proof of (P*)? I am going to answer both questions in the negative, on the condition of a certain assumption. The assumption is that each level in the hierarchy is taken to come with a decidable set of axioms and rules, so that it can be represented as a formal system. In order to see the problem, consider gödel sentences in *L*. It is assumed that *L* contains the language and standard axioms of arithmetic. Under these assumptions, which I think Prawitz does make, there are gödel sentences in *L*. Suppose that $S = (\forall xGx)^{\circ}$ is a gödel sentence in *L*, given some gödel numbering. Then a proof of *S* will as usual involve the observation that every instance is provable by the proof rules in *L*, and this reasoning can be carried out in *ML*. That is, we will have as a truth in *ML*

$$\forall t^{\circ}\exists a(P(a, (Gt)^{\circ})) \tag{9}$$

We may suppose that the proof rules of *ML* allow the derivation of (9). We may also assume that the proof rules of *ML* allow the (correct) inference from (9) to

$$\forall xGx \tag{10}$$

i.e. to *S* itself. This can be stated in *ML*, since *ML* is an extension of *L*.

It might now seem, since we have a proof of *S*, that we can go on to claim that there *is* a proof of *S*, i.e. to claim

$$\exists a(P(a, (\forall x Gx)^{\circ})) \tag{11}$$

This would be a mistake, however. The proof variable 'a' in (11) ranges over proofs in the *ML* domain of proofs, and these are proofs constructible *in L*. The proof of S carried out in *ML* is not, however, constructible in L, precisely since S is a gödel sentence for L. Therefore, (11) is false: there is (on the assumption that the set of axioms and proof rules for L is consistent) no element in the domain of discourse of *ML* that is a proof of S.

By contrast, we *can* correctly say in *MML* that there is a proof of S, since the proof variables of *MML* range over proofs constructible in *ML* (as well as in L). That is, we can correctly say in *MML*

$$\exists \boldsymbol{a}(\boldsymbol{P}(\boldsymbol{a}, (\forall x Gx)^{\bullet})) \tag{12}$$

But now, the fact that (12) is assertible but (11) is not, has negative consequences for (P*). For consider the instance of (P*)

$$\exists \boldsymbol{b}(\boldsymbol{P}(\boldsymbol{b}, (\forall x Gx \to \exists a(P(a, (\forall x Gx)^{\circ})))^{\bullet})) \tag{13}$$

(13) is true if there is an element in the *MML* domain of discourse that is a proof of the *ML* sentence

$$\forall x Gx \to \exists a(P(a, (\forall x Gx)^{\circ})) \tag{14}$$

Such a proof $\boldsymbol{b}$ is to be a function that maps any proof $\boldsymbol{a}$ of the antecedent, i.e. S, in the *MML* domain, on a proof $\boldsymbol{b}(\boldsymbol{a})$ of the consequent, i.e. $\exists a(P(a, (\forall x Gx)^{\circ}))$, again in the *MML* domain. But there is no such proof $\boldsymbol{b}$, because there is a proof of the antecedent in the domain but no proof (at all) of the consequent. Hence, (13) is false. It immediately follows that (P*) itself is false, since (13) is an instance of (P*).

The argument above, if correct, constitutes a refutation of (P*). On the other hand, Prawitz has provided a sketch of a proof of (P*). So what has gone wrong? The crucial part of Prawitz's induction steps are those corresponding to the equating of (5) and (6). In the case of the universal quantifier, the corresponding part would be a principle, derived from the intuitionistic meaning explanation of the universal quantifier, saying that a canonical proof of a sentence

$$P(c, (\forall x Ax)^{\circ}) \tag{15}$$

of *ML* is the same as a canonical proof of

$$\forall t^{\circ}(P((\textsc{run}(c))(t^{\circ}), (At)^{\circ})) \tag{16}$$

where 't°' ranges over closed singular terms of L. This is to say that a canonical proof that c is a proof of $(\forall x Ax)^{\circ}$ is the same as a canonical proof that the canonical proof extractable from c, when applied to a closed term t° of L, yields a proof of $(At)^\circ$.

The problem with equating (15) and (16) is that this constitutes a *reflection principle* for L itself. For suppose we have a canonical proof of (16). Then, again, we have shown that there is a certain construction c which when applied to a term t° of L yields a proof of $(At)^\circ$. But now, if the predicate A happens to be the gödel predicate G, then there is indeed, for every term t° a proof $c(t^\circ)$ of $(Gt)^\circ$, but still no proof in the domain of *ML*, i.e. constructible in L, of $(\forall x Gx)^\circ$. Yet, if a canonical proof of the gödel instance of (16) also is a canonical proof of the corresponding gödel instance of (15), there is, contrary to assumption, a proof in the domain of *ML* of the gödel sentence.[13]

The situation is obscured by the natural assumption that if '$c(t^\circ)$' denotes a proof constructible in L, for any L numeral t°, then so does 'c' itself (or '$\lambda x(c(x))$'). But this cannot be right, for it seems that '$c(t^\circ)$' does denote a proof, although not a canonical proof, of $(Gt)^\circ$ in L, for any numeral t°. For the gödel predicate 'G' is a decidable predicate, which means that it can be proved in *ML* that

$$\exists a(P(a, (\forall x(Gx \vee \neg Gx))^\circ)) \tag{17}$$

The proof a is a method for getting a canonical proof, which in turn is a function which, when applied to a term t° yields a proof of the corresponding instance. That is

$$\exists a(\forall t^\circ(P((\textsc{run}(a))(t^\circ), (Gt \vee \neg Gt)^\circ))) \tag{18}$$

Suppose that e is such a proof, and u° a term. Then we have

$$P((\textsc{run}(e))(u^\circ), (Gu \vee \neg Gu)^\circ) \tag{19}$$

but $\textsc{run}(e))(u^\circ)$ is not a canonical proof. The canonical proof of $(Gu \vee \neg Gu)^\circ$ introduces the disjunction from either a proof of $(Gu)^\circ$ or a proof of $(\neg Gu)^\circ$. As usual, the canonical proof can be extracted from the non-canonical, or indirect, proof (in this case by iterating instances of the proof of the induction step in u). That is, it is provable in *ML* that

$$P(\textsc{run}((\textsc{run}(e))(u^\circ)), (Gu)^\circ) \vee P(\textsc{run}((\textsc{run}(e))(u^\circ)), (\neg Gu)^\circ) \tag{20}$$

This generalizes into

[13] Prawitz [12, p. 318] draws attention to the realizability formula '$A \leftrightarrow \textcircled{R} A$' of [8]. As far as I understand, beside the difference Prawitz himself points to, there is a further difference that is crucial. The counterparts to (15) and (16) in [8] *are* equivalent, simply because the counterpart to (15) is defined to be an abbreviation of the counterpart to (16). This means that, in primitive notation, '$\forall x Ax$' does not even occur as a proper sub-expression. Hence, there is no substantial issue of equating or not.

$$\forall t^\circ(P(\textsc{run}((\textsc{run}(e))(t^\circ)), (Gt)^\circ) \vee P(\textsc{run}((\textsc{run}(e))(t^\circ)), (\neg Gt)^\circ)) \tag{21}$$

Now, the proof that there is a proof of $(Gt)^\circ$ for any numerical term t° makes use of the construction of G° and theorems such as the Diagonal Lemma, the proof of which is not available in *L* itself. By these means it can be proved, in *ML*, that the assumption

$$\exists a(P(a, (\exists x(\neg Gx))^\circ)) \tag{22}$$

leads to a contradiction. From this it is quickly inferred in *ML* that

$$\forall t^\circ(\neg P(\textsc{run}((\textsc{run}(e))(t^\circ)), (\neg Gt)^\circ)) \tag{23}$$

And from (23) and (21) it can be proved in *ML* that

$$\forall t^\circ(P(\textsc{run}((\textsc{run}(e))(t^\circ)), (Gt)^\circ)) \tag{24}$$

By the extrapolated Prawitz reflection principle, i.e. the equating of (15) and (16), a canonical proof in *ML* of (24) is also a canonical proof of

$$P(\lambda t^\circ(\textsc{run}((\textsc{run}(e))(t^\circ))), (\forall x Gx)^\circ) \tag{25}$$

(25), however, is false. The lambda term '$\lambda t^\circ(\textsc{run}((\textsc{run}(e))(t^\circ)))$' does not denote any proof constructible in *L*. This would be a free variable construction in *L*, giving a proof of $(Gt)^\circ$ for arbitrary term t°. There is no such thing. Instead, there is a different construction for each term, even though each of these constructions can be extracted from the induction proof of $(\forall x(Gx \vee \neg Gx))^\circ$, given a term t°. That this is the case, again, is proved by a free variable construction in *ML*, in the manner hinted above. This free variable construction is, in turn, denoted by a lambda term in *MML*.

This refutation of the reflection principle, and thus of Prawitz's induction proof of (P*) has proceeded on the assumption that the set of rules and axioms of *L* is decidable, and therefore can be regarded as a formal system. The same was assumed of each level in the hierarchy, although the total hierarchy itself can not be seen as a formal system. But suppose we give up that assumption about *L*. Will Prawitz's construction then go through?

5 Consequences of Prawitz's Construction

Giving up the assumption that *L* together with it axioms and rules of inference is a formal system allows us to accept the reflection principle involved in the universal quantifier step of Prawitz's induction proof of (P*), and then I can see no reason to think that the induction proof will not go through at this level, i.e. for sentences of *L*. I believe it will.

If we don't treat *L* as a formal system we can regard the lambda term '$\lambda t^{\circ}(\text{RUN}((\text{RUN}(e))(Gt)^{\circ})))$' as denoting a proof in *L*, e.g. a separate axiom. Similarly, for each sentence *S* of *L* of which it can be proved in *ML* that it is a gödel sentence for the original set of rules, *S* is an axiom in the new, expanded set of axioms. The expanded set of axioms is not recursive, but it is still recursively enumerable, since each axiom corresponds to a theorem provable in *ML*, assuming that *ML*, with *its* rules and axioms, still is a formal system.

If the set of rules and axioms of *ML* is decidable, *ML* can still be treated as a formal system, and then has its own incompleteness and its own gödel sentences. Because of this, however, we will again have a counter-example to the Prawitz induction proof when we want to move it one level up, and state in *MMML* that for each sentence of *ML* there is a proof constructible in *MML*. In order to eliminate the counter-example at this level, we need to accept a reflection principle for *ML*. The rules for *ML* will therefore not make a formal system, either. And, as far as I can see, to make the induction proof go through in full generality, i.e. for all levels, the same must be done at each level. Therefore *no* level in the linguistic hierarchy comes with a decidable set of rules.

But this has consequences already at *L*. For among the new axioms of *ML* that are added because of the reflection principle for *ML*, are sentences about proofs in *L*, and this in turn induces even more new axioms for *L*, because of the reflection principle for *L*. There will then also be even more new axioms for *ML* and for *L* because of the reflection principle for *MML*, and so on. So, a sentence is an axiom in *L* if it is a consequence of what is provable at *some level or other* in the hierarchy. But since no level will have a decidable set of rules and axioms, this means that the set of axioms of *L* will not even be recursively enumerable.

As a consequence, it is not clear what remains of the concept of a proof over and above the concept of being true in a non-constructive sense. If in effect we say that any sentence that is *true* is to be added as an axiom of *L*, then provability is simply reduced to truth. This seems to be the ultimate consequence: by availing oneself of the means to prove (P*) in its full generality, if only level by level, provability reduces to truth in a non-constructive sense. Or at least, it is not clear to me how this consequence can be avoided. Of course, because the equivalence principle (EP) holds for 'true', (P*) is clearly correct in case of such a reduction.

6 Conclusion

Since there is reason to think that we cannot, for *every* sentence *A*, come to know either that *A* is true or that *A* is false, we also have a reason to doubt the principle of bivalence, *provided* there is reason to think that

If a sentence *A* is true, it is also provable/knowable.

This is intuitively what (P*) says. In the absence of a reason for holding on to (P*), or at least an interesting restricted version of it, there is no reason to doubt bivalence either. That is, we have been given no justification of the claim that it isn't valid.

What does hold, if the reasons for intuitionistic semantics are correct, is that bivalence does *not* have the status of a quasi-logical principle. In a classical semantics it does, since it is justifiable from the law of excluded middle together with the equivalence principle (EP), and the equation of falsity with truth of negation. But that it isn't a principle that enjoys the status of quasi-logical validity is not a reason for thinking that it isn't valid. In fact, it is precisely because of this that Dummett's connection between bivalence and realism is to the point. Belief in bivalence is an expression of realism because bivalence is *not* anti-justifiable by meaning theory and logic alone.

References

1. Dummett, M., 1976, 'What is a theory of meaning? (II)', in G. Evans and J. McDowell (eds.), *Truth and Meaning*, Oxford University Press, Oxford. Reprinted in Dummett 1998, 34–93. Page references to the reprint.
2. Dummett, M., 1980, 'The philosophical significance of Gödel's theorem', in *Truth and Other Enigmas*, 186–201. Duckworth, London, 2nd edn., Originally published in *Ratio* V, 1963, 140–155.
3. Dummett, M., 1982, 'Realism', *Synthese* 52:55–112.
4. Dummett, M., 1991, *The Logical Basis of Metaphysics*, Harvard University Press, Cambridge, Mass.
5. Dummett, M., 1998, *The Seas of Language*, Clarendon Press, Oxford.
6. Kreisel, G., 1962, 'Foundations of intuitionistic logic', in E. Nagel (ed.), *Logic, Methodology and Philosophy of Science I*, North-Holland, Amsterdam.
7. Martin-Löf, P., 1984, *Intuitionistic Type Theory*, Bibliopolis, Napoli.
8. Nelson, D., 1947, 'Recursive functions and intuitionistic number theory', *Transactions of the American Mathematical Society* 61:307–368.
9. Pagin, P., 1998, 'Bivalence: meaning theory vs metaphysics', *Theoria* LXIV:157–186.
10. Prawitz, D., 1980, 'Intuitionistic logic: a philosophical challenge', in G. H. von Wright (ed.), *Logic and Philosophy*, Martinus Nijhoff Publishers, The Hague.
11. Prawitz, D., 1994, 'Meaning theory and anti-realism', in B. McGuinnes and G. Oliveri (eds.), *The Philosophy of Michael Dummett*, Kluwer, Dordrecht, 79–89.
12. Prawitz, D., 1998, 'Comments on Peter Pagin's paper', *Theoria* LXIV.
13. Tennant, N., 1996, 'The law of excluded middle is synthetic a priori, if valid', *Philosophical Topics* 24:205–229.

From Intuitionistic to Point-Free Topology: On the Foundation of Homotopy Theory

Erik Palmgren*

1 Introduction

Brouwer's pioneering results in topology, e.g. invariance of dimension, were developed within a classical framework of mathematics. Some years later he explained that "in his topological work he tried to use only methods which he expected could be made constructive" [10, p. XIV]. It seems that very little of algebraic topology, homotopy and homology, has actually been developed constructively in any detail, or, at any rate, found its way to publication. In the comprehensive treatise *Foundations of Constructive Mathematics,* Beeson [4, pp. 26 – 27] writes however

> The classical results of algebraic topology do not require the general concept of a topological space. If we content ourselves to treat metric spaces, then the standard treatments of the homotopy and homology groups are quite straightforwardly constructive, e.g. Greenberg [1967][17]. One draws all the usual corollaries, e.g. $\mathbb{R}^n$ and $\mathbb{R}^m$ are not homeomorphic unless $n = m$; [...] It is quite essential to deal with uniformly continuous functions, and not just with continuous functions.

To be able to make certain quotient and glueing constructions it is necessary to have a constructive theory of more general topological spaces than metric spaces. As argued by many authors, locales, or point-free topologies, should provide a good constructive foundation for topological theories; see [2, 8, 3, 18, 19]. The point-free approach can also be regarded as a (re)interpretation of Brouwer's theory of choice sequences (Martin-Löf [17], Coquand [6, p. 31]), whereby one regains covering compactness results, e.g. the Heine-Borel theorem, but with a fully constructive interpretation. For some background and history of point-free topology from lattice-theoretic origins, see [13, 14]. It is interesting to note that such ideas were also introduced in the intuitionistic school, first by Freudenthal [9], and then further developed by Troelstra [22]. Already in [9], which treats compact Hausdorff spaces, the characteristics of point-free topology were clearly visible: some class

E. Palmgren (✉)
Department of Mathematics, Uppsala University, Uppsala, Sweden
e-mail: palmgren@math.uu.se; URL: www.math.uu.se

*The author is supported by a grant from the Swedish Research Council (VR).

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_12, © Springer Science+Business Media B.V. 2009

of open sets form the fundamental objects, points are merely derived and covering relations are defined without reference to points. There is also a development of general topology [23] using non-effective principles, like Brouwer's Fan Theorem (which is valid in certain toposes).

Fundamental groups for locales have been constructed by W. He [12]. Various generalisations and equivalents of fundamental groups for locales and toposes are studied by Kennison [16]. Most prominently a point-free approach to paths and homotopy is developed. These papers are not explicitly concerned with constructivity, and certainly not predicativity. We shall here investigate a key element in the construction of the fundamental group for a formal topology, combining those foundational perspectives. Formal topology is a predicative version of locale theory, due to Martin-Löf and Sambin; see [21]. We shall therefore work within the framework of constructive mathematics in the style of Bishop (cf. [5]), which in turn can be formalised in constructive type theory, or constructive set theory. In fact some of the results, Theorems 8 and 17, do not use any kind of choice principle and are therefore valid in any topos (cf. [15]).

2 Paths in Spaces

In the construction of the fundamental group of a space, joining and deformation of paths in a space are basic operations.

A space Y is said to have the *path joining property (PJP)* if $f\colon [a,b] \longrightarrow Y$ and $g\colon [b,c] \longrightarrow Y$ are continuous functions with $f(b) = g(b)$, then there exists a unique continuous function $h\colon [a,c] \longrightarrow Y$, with $h(t) = f(t)$ for $t \in [a,b]$, and $h(t) = g(t)$ for $t \in [b,c]$. Classically, every topological space satisfies PJP since we can define h by cases and then check its continuity. Constructively, case-wise definition is a priori not possible. Furthermore, uniform continuity, in some form, is the basic notion. In this section, the discussion is restricted to metric spaces.

Lemma 1 *Let X be a metric space and let Y be a complete metric space. If $D \subseteq X$ is a dense subset, and $f\colon D \longrightarrow Y$ is uniformly continuous, then there is a unique continuous function $h\colon X \longrightarrow Y$ so that $h(t) = f(t)$ for all $t \in D$.*

Proof The existence follows from [5, Lemma 3.3.7]. Uniqueness is direct from continuity, using the inequality $|h_1(x)-h_2(x)| \leqslant |h_1(x)-h_1(t)|+|h_2(t)-h_2(x)|$, where $t \in D$. □

A dense set in the interval $[a,b]$ is, for instance, $D_{a,b} = \{a + (b-a)k2^{-n} : n = 1,2,3,\ldots; k = 0,\ldots,2^n\}$.

Theorem 2 *PJP is valid for complete metric spaces Y.*

Proof Let $f\colon [a,b] \to Y$ and $g\colon [b,c] \to Y$ be continuous functions. The set $E = D_{a,b} \cup D_{b,c}$ is dense in $[a,c]$. Since $D_{a,b} \cap D_{b,c}$ contains only the element b, and $f(b) = g(b)$, the following is indeed a definition of a function $k : E \to Y$. Let $k(t) = y$ if, and only if,

$$t \in D_{a,b} \text{ and } f(t) = y, \text{ or } t \in D_{b,c} \text{ and } g(t) = y.$$

If ω_f and ω_g are the continuity moduli of f and g, respectively, then

$$\omega_k(\varepsilon) = {}_{\mathrm{def}}\min\left(\omega_f\left(\frac{\varepsilon}{2}\right), \omega_g\left(\frac{\varepsilon}{2}\right)\right)$$

is a modulus for k. Let h be the unique extension of k according to Lemma 1. Again by this lemma, h restricted to $[a, b]$ is f, and h restricted to $[b, c]$ is g. It is also the unique such h. □

This method fails for general metric spaces.

Proposition 3 *There is a metric space Y, such that if the PJP is valid for Y, then for any real x*

$$x \leqslant 0 \textit{ or } x \geqslant 0.$$

Proof Let Y be $[-1, 0] \cup [0, 1]$ as metric subspace of $\mathbb{R}$, and suppose that PJP is valid for Y. Let $f\colon [-1, 0] \longrightarrow Y$ and $g\colon [0, 1] \longrightarrow Y$ be inclusion maps. Take $h\colon [-1, 1] \longrightarrow Y$ to be the unique common extension of those maps, as given by PJP. Thus $h_{\restriction[-1,0]} = f$ and $h_{\restriction[0,1]} = g$. We show that $h(x) = x$ for all $x \in [-1, 1]$. Consider the inclusion $k\colon Y \longrightarrow [-1, 1]$, and the composition $k \circ h\colon [-1, 1] \longrightarrow [-1, 1]$. Now $(k \circ h)_{\restriction[-1,0]} = k \circ f$, i.e. the inclusion of $[-1, 0]$ into $[-1, 1]$, and $(k \circ h)_{\restriction[0,1]} = k \circ g$, i.e. the inclusion of $[0, 1]$ into $[-1, 1]$. The identity map $\mathrm{id}_{[-1,1]}$ restricted similarly, gives obviously the same inclusions. By Theorem 2 for $[-1, 1]$, we get $k \circ h = id_{[-1,1]}$. Thus $h(x) = k(h(x)) = x$ for all $x \in [-1, 1]$. But $h(x) \in Y$, so for all $x \in [-1, 1]$,

$$x \leqslant 0 \text{ or } 0 \leqslant x.$$

The desired conclusion follows, by noting that for any real x it holds that $x \in [-1, 1]$, $x \leqslant -1/2$ or $1/2 \leqslant x$. □

As the conclusion is non-constructive [5], the proposition provides a Brouwerian counterexample to the general validity of PJP.

A continuous deformation of a function into another function is a homotopy [1]. Composition of such two deformations can be regarded as generalised joining operation: It takes two continuous functions $f\colon X \times [a, b] \longrightarrow Y$ and $g\colon X \times [b, c] \longrightarrow Y$ which agree at the edges b as: $f(x, b) = g(x, b)$ for all $x \in X$. The operation should then join them by producing a continuous function $h\colon X \times [a, c] \longrightarrow Y$ so that h agrees with f and g on $X \times [a, b]$ respectively $X \times [b, c]$. It can be expressed in terms of maps only using a commutative diagram as follows. (This formulation will be useful in Section 3.) A pair of spaces (X, Y) has the *homotopy joining property* (HJP), if the diagram

$$\begin{tikzcd} X \arrow[r, "\langle 1_X, \hat{b}\rangle"] \arrow[d, "\langle 1_X, \hat{b}\rangle"'] & X \times [b,c] \arrow[d, "1_X \times j_2"] \\ X \times [a,b] \arrow[r, "1_X \times j_1"'] & X \times [a,c] \end{tikzcd} \tag{1}$$

commutes, and for any uniformly continuous $F: X \times [a, b] \longrightarrow Y$ and $G: X \times [b, c] \longrightarrow Y$ with $F \circ \langle 1_X, \hat{b}\rangle = G \circ \langle 1_X, \hat{b}\rangle$, there exists a unique uniformly continuous $H: X \times [a, c] \longrightarrow Y$ so that $F = H \circ (1_X \times j_1)$ and $G = H \circ (1_X \times j_2)$. Here the $\hat{b}$'s are constant maps with value b on X, and the j's are inclusions of intervals. (The case where $X = 1$ (the one point metric space) is PJP for Y.)

Theorem 4 *HJP holds for (X,Y) when X is a metric space and Y is a complete metric space.*

Proof Analogous to Theorem 2, using the dense set $X \times E$ in $X \times [a, c]$. □

Using this theorem it is possible to construct the fundamental group at a point of a complete metric space along standard lines [1]. Only the cases where X is a finite, closed interval are needed. HJP is used, in particular, when showing that the homotopy relation is transitive: Suppose that p and q are homotopic, and that q and r are homotopic. Then there are uniformly continuous $F{:}X \times [0, 1] \longrightarrow Y$ and $G: X \times [1, 2] \longrightarrow Y$ so that for all $x \in X$

$$F(x, 0) = p(x) \quad F(x, 1) = q(x) = G(x, 1) \quad G(x, 2) = r(x).$$

Then HJP gives the desired homotopy $H: X \times [0, 2] \longrightarrow Y$ deforming p into r.

3 Formal Topology

In contrast to the negative result of the previous section, the category of formal topologies satisfies HJP for all pairs of spaces (X, Y). Equivalently, it can be stated as: the diagram (1) is a pushout in this category, for every object X. This main result in shown in Section 4. In Section 3.1 we review of some of the basics of formal topology. Sections 3.2, 3.3, 3.4 prepares for the main result.

3.1 Basic Definitions and Results

Definition 5 A formal topology consists of a pre-order $X = (X, \leqslant)$ of basic open neighbourhoods and $\lhd \subseteq X \times \mathcal{P}(X)$, the covering relation, satisfying the four covering conditions

(Ref) $a \in U \Longrightarrow a \lhd U$,
(Tra) $a \lhd U, U \lhd V \Longrightarrow a \lhd V$,
(Loc) $a \lhd U, a \lhd V \Longrightarrow a \lhd U_{\leqslant} \cap V_{\leqslant}$,
(Ext) $a \leqslant b \Longrightarrow a \lhd \{b\}$.

Here $U \lhd V \Leftrightarrow_{\text{def}} (\forall a \in U)\, a \lhd V$, and moreover, $Z_{\leqslant} =_{\text{def}} \{x \in X\colon (\exists z \in X)\ x \leqslant z\}$ is the down-closure of Z. Furthermore we require that the cover relation is set-presented, in the sense that there is a family $\{C(a, i)\}_{i \in I(a)}$ of subsets of X so that

$$a \lhd U \Longleftrightarrow (\exists i \in I(a))\, C(a, i) \subseteq U.$$

We write the components of a formal space $\mathcal{X}$ as $(X, \leqslant_{\mathcal{X}}, \lhd_{\mathcal{X}}, C_{\mathcal{X}})$, often omitting the set-presentation $C_{\mathcal{X}}$.

Define the mutual cover relation $U \sim V$ to hold iff $U \lhd V$ and $V \lhd U$. Let $Z_{\lhd} = \{x \in X : x \lhd Z\}$. A subset $Z \subseteq X$ is *saturated* if $Z_{\lhd} = Z$. The saturated subsets corresponds to elements in the associated locale. They may always be represented by subsets up to mutual covering, since $U \sim U_{\lhd}$. Any subset represents an open set in this way. A subset $Z \subseteq X$ is *down-closed* if $Z_{\leqslant} = Z$.

A pair (a, U), where $a \in X$ and $U \subseteq X$, is called a *covering axiom*. A formal topology $\mathcal{X}$ is *generated by a family of covering axioms* (a_i, U_i) $(i \in I)$, if $\lhd_{\mathcal{X}}$ is the smallest relation satisfying covering conditions and the axioms

$$a_i \lhd_{\mathcal{X}} U_i \quad (i \in I).$$

From the set-presentation of a formal topology $\mathcal{X}$ one can obtain a generating family of covering axioms $(b_j, V_j)_{j \in J}$ for $\mathcal{X}$ as follows. Let $J = \{(a, i) : a \in X, i \in I_{\mathcal{X}}(a)\}$ and put $b_{(a,i)} = a,\ V_{(a,i)} = C_{\mathcal{X}}(a, i)$. Conversely, one can show (see e.g. [20]) that if $\mathcal{X}$ satisfies axioms (Ref), (Tra), (Loc) and (Ext) and is generated by a set-indexed family of covering axioms, then $\mathcal{X}$ is a formal topology. It is often easier to exhibit a set of covering axioms, than a set-presentation.

3.1.1 Points

A *point* of $\mathcal{X}$ is a non-void subset $\alpha \subseteq S$ which is

(Fil) $\leqslant$-filtering, i.e. for $a, b \in \alpha$, there is $c \in \alpha$ with $c \leqslant a$ and $c \leqslant b$,
(Spl) such that α contains a neighbourhood from U, whenever $a \lhd U$ and $a \in \alpha$. (This is often expressed as: "a point splits any cover").

The points of a formal topology $\mathcal{X}$ form a class $\text{Pt}(\mathcal{X})$, which, under certain conditions, is a set. For $a \in X$, let a^* denote the subclass of points in $\mathcal{X}$ satisfying $a \in \alpha$. For a subset $U \subseteq X$, let U^* denote the union of all the subclasses a^* for $a \in U$.

Lemma 6 *Any formal cover in $\mathcal{X}$ is a point-wise cover:*

$$a \lhd_{\mathcal{X}} U \Longrightarrow a^* \subseteq \cup U^*. \quad \square$$

The converse implication is not true in general, and this explains why results like the Heine-Borel Theorem are possible to prove in this setting. We say the covers of formal topology $\mathcal{X}$ are *order conservative,* if $a \leqslant_{\mathcal{X}} b$ whenever $a \lhd_{\mathcal{X}} \{b\}$. This notion is of course only interesting if $\leqslant$ has a simpler definition than $\lhd$. (Any formal topology is isomorphic to an order conservative one where $\lhd$ is the partial order.) The covers are *point-wise order conservative* if $a^* \subseteq b^*$ implies $a \leqslant_{\mathcal{X}} b$. In view of Lemma 6 the latter is a stronger property.

3.1.2 Continuous Morphisms

Let $\mathcal{S} = (S, \leqslant, \lhd)$ and $\mathcal{T} = (T, \leqslant', \lhd')$ be formal topologies. A relation $F \subseteq S \times T$ is a *continuous mapping* $\mathcal{S} \longrightarrow \mathcal{T}$ if

(A1) $a\,F\,b, b \lhd' V \Longrightarrow a \lhd F^{-1}\,V$,
(A2) $a \lhd U, x\,F\,b$ for all $x \in U \Longrightarrow a\,F\,b$,
(A3) $S \lhd F^{-1}\,T$,
(A4) $a\,F\,b, a\,F\,c \Longrightarrow a \lhd F^{-1}(b_{\leqslant'} \cap c_{\leqslant'})$.

Here $F^{-1}Z = \{x \in S : (\exists y \in Z)\, x R y\}$ and $z_{\leqslant'}$ is $\{z\}_{\leqslant'}$. It is possible to replace the quantifications over the subsets U and V, with quantification over the set-presentations of $\mathcal{S}$ and $\mathcal{T}$ respectively.

Some equivalent versions of the above axioms are

(A1') $b \lhd' V \Longrightarrow F^{-1}b \lhd F^{-1}\,V$,
(A2') $a \lhd F^{-1}b \Longrightarrow a\,F\,b$,
(A4') $F^{-1}U \cap F^{-1}V \lhd F^{-1}(U_{\leqslant'} \cap V_{\leqslant'})$.

We have for any continuous F that $F^{-1}U \lhd F^{-1}V$, if $U \lhd' V$. Hence $F^{-1}U \sim F^{-1}V$ whenever $U \sim' V$. Also by (A1) $F^{-1}(U_{\lhd'}) \sim F^{-1}U$. By (A2) it follows that each $F^{-1}Z$ is down-closed.

Each continuous mapping induces a point function $f = \mathrm{Pt}(F)$ given by

$$\alpha \mapsto \{b : (\exists a \in \alpha) F(a, b)\} : \mathrm{Pt}(\mathcal{S}) \longrightarrow \mathrm{Pt}(\mathcal{T})$$

and which satisfies: $a\,F\,b \Rightarrow f[a^*] \subseteq b^*$.

Composition of two continuous morphisms F: $\mathcal{X} \longrightarrow \mathcal{Y}$ and G: $\mathcal{Y} \longrightarrow \mathcal{Z}$ is given as follows

$$a\,(G \circ F)\,c \Longleftrightarrow a \lhd_{\mathcal{X}} F^{-1}[G^{-1}(c)].$$

The one-point formal topology is the terminal object in the category of formal topologies. It is constructed as $\mathbf{1} = (\{*\}, \leqslant_{\mathbf{1}}, \lhd_{\mathbf{1}})$, where $* \leqslant_{\mathbf{1}} *$ and $a \lhd_{\mathbf{1}} U$ iff U is inhabited. The terminal map $!_{\mathcal{Y}}$ from $\mathcal{Y}$ to $\mathbf{1}$ is defined by letting the relation $y\ !_{\mathcal{Y}}\ a$ be true for all y and a.

Now any point $\alpha \in \mathrm{Pt}(\mathcal{X})$ in a formal topology, gives a unique morphism $F_\alpha : \mathbf{1} \longrightarrow \mathcal{X}$, which is given by

$$a \, F_\alpha \, x \Longleftrightarrow x \in \alpha.$$

A map $\hat{\alpha} : \mathcal{Z} \longrightarrow \mathcal{X}$ which is constant α is defined by the composition $F_\alpha \circ !_{\mathcal{Z}}$. More explicitly, the map is given by the relation

$$z \, \hat{\alpha} \, x \Longleftrightarrow z \lhd_{\mathcal{Z}} \{u \in Z : x \in \alpha\}.$$

In particular, if z is covered by the empty set, then $z \, \hat{\alpha} \, x$ holds for any x.

3.2 Closed Subspaces

Let $\mathcal{X} = (X, \leqslant, \lhd)$ be a formal topology. A subset $U \subseteq X$ defines an open set in the topology. It also defines a *closed subspace* by its formal complement as follows. Let $\mathcal{X} \dot{-} U = (X, \leqslant, \lhd')$ where

$$a \lhd' V \Longleftrightarrow a \lhd U \cup V.$$

(Note that $\lhd'$ is generated by the covering axioms for $\lhd$ and the pairs $(a, \emptyset)$ for $a \lhd U$.) By the definition of $\lhd'$ we see that

$$X \dot{-} U = X \dot{-} (U_\lhd). \tag{2}$$

Proposition 7 *Let $\mathcal{X}$ be a formal topology. For $S \subseteq X$, we have*

$$\alpha \in \mathrm{Pt}(\mathcal{X} \dot{-} S) \Longleftrightarrow \alpha \in \mathrm{Pt}(\mathcal{X}) \textit{ and } \alpha \notin S^*.$$

We shall consider inclusion mappings between closed subspaces of a formal topology $\mathcal{X}$. For subsets $V \subseteq U \subseteq X$, let $E_{U,V} : \mathcal{X} \dot{-} U \longrightarrow \mathcal{X} \dot{-} V$ be defined by

$$x \, E_{U,V} \, y \Longleftrightarrow_{\mathrm{def}} x \lhd_{(\mathcal{X} \dot{-} U)} \{y\}.$$

The right hand side is thus equivalent to $x \lhd_{\mathcal{X}} U \cup \{y\}$, and hence we have

$$a \lhd_{\mathcal{X} \dot{-} U} E_{U,V}^{-1} W \Longleftrightarrow a \lhd_{\mathcal{X}} U \cup W. \tag{3}$$

Each morphism $E_{U,V}$ is a monomorphism in the category of formal topologies. Furthermore it follows that

$$E_{V,W} \circ E_{U,V} = E_{U,W} \tag{4}$$

for $W \subseteq V \subseteq U \subseteq X$. We shall write E_U for $E_{U,\emptyset} : (\mathcal{X} \dot{-} U) \longrightarrow \mathcal{X}$. Note that

$$E_{U,V} = E_{U_\lhd, V_\lhd}. \tag{5}$$

3.2.1 A Glueing Theorem

We generalise slightly the glueing theorem from [18] (there stated without proof).

Theorem 8 *Let $\mathcal{X}$ be a formal topology. Let I be a finite index set, and suppose that $U_i \subseteq X$ is down-closed for each $i \in I$. Suppose that $F_i : (\mathcal{X} \dot{-} U_i) \longrightarrow \mathcal{Y}$, $i \in I$, are continuous morphisms such that for all $i, j \in I$,*

$$F_i \circ E_{U_{ij},U_i} = F_j \circ E_{U_{ij},U_j} \tag{6}$$

where $U_{ij} = U_i \cup U_j$. Let $W = \bigcap_{i \in I} U_i$. Then there is a unique $F \colon (\mathcal{X} \dot{-} W) \longrightarrow \mathcal{Y}$ such that

$$F \circ E_{U_i,W} = F_i \tag{7}$$

for all $i \in I$.

Proof Note that by definition of composition and (3) we have

$$\begin{aligned} a\,(F_i \circ E_{U_{ij},U_i})\,b \quad &\Longleftrightarrow \quad a \lhd_{\mathcal{X} \dot{-} U_{ij}} E^{-1}_{U_{ij},U_i} F_i^{-1} b \\ &\Longleftrightarrow \quad a \lhd_{\mathcal{X}} U_{ij} \cup F_i^{-1} b \\ &\Longleftrightarrow \quad a \lhd_{\mathcal{X}} U_i \cup U_j \cup F_i^{-1} b \\ &\Longleftrightarrow \quad a \lhd_{\mathcal{X}} U_j \cup F_i^{-1} b. \end{aligned}$$

The last step follows since $U_i \subseteq F_i^{-1} b$. By a similar equivalence for $F_j \circ E_{U_{ij},U_j}$, the equation (6) can be read as the equivalence: for all a and b,

$$a \lhd_{\mathcal{X}} U_j \cup F_i^{-1} b \Longleftrightarrow a \lhd_{\mathcal{X}} U_i \cup F_j^{-1} b. \tag{8}$$

Any saturated set, such as $F_i^{-1} b$, is down-closed as well. This property is used frequently when applying the localisation axiom (Loc).

We show uniqueness first, which gives an explicit definition of F. Suppose that F satisfies (7). Then for any $i \in I$, by expanding definitions and using transitivity,

$$a \lhd_{\mathcal{X}} U_i \cup F^{-1} b \Longleftrightarrow a \lhd_{\mathcal{X}} U_i \cup E^{-1}_{U_i,W} F^{-1} b \Longleftrightarrow a\, F_i\, b. \tag{9}$$

We use localisation to obtain

$$a \lhd_{\mathcal{X}} \bigcap_{i \in I} (U_i \cup F^{-1} b) \Longleftrightarrow (\forall i \in I)\; a\, F_i\, b. \tag{10}$$

Note that finiteness of I is essential here. Now

$$\bigcap_{i\in I}(U_i \cup F^{-1}b) = (\bigcap_{i\in I} U_i) \cup F^{-1}b = W \cup F^{-1}b,$$

so the left hand side of (10) is equivalent to $a\ F\ b$. Thus we have

$$a\ F\ b \Longleftrightarrow (\forall i \in I)\ a\ F_i\ b$$

which is an explicit definition of F, which is therefore uniquely determined.

Suppose now that F is given by this explicit definition. We show that it satisfies (7) and is continuous. To show (7), note first that we have by the definition of $\circ$

$$a\ F \circ E_{U_i,W}\ b \Longleftrightarrow a \lhd_{\mathcal{X}} U_i \cup F^{-1}b. \tag{11}$$

Since $F \subseteq F_i$, its right hand side implies $a \lhd_{\mathcal{X}} U_i \cup F_i^{-1}b$, that is $a\ F_i\ b$. Thus we have $F \circ E_{U_i,W} \subseteq F_i$. To show the reverse inclusion, suppose that $a\ F_i\ b$. From (8) we get, for any j, that $a \lhd_{\mathcal{X}} U_i \cup F_j^{-1}b$. Then applying localisation,

$$a \lhd_{\mathcal{X}} U_i \cup \bigcap_{j\in I} F_j^{-1}b.$$

But $\bigcap_{j\in I} F_j^{-1}b = F^{-1}b$, by definition of F, so this gives, via (11), $a\ F \circ E_{U_i,W}\ b$. This proves (7).

Finally, we show that F is continuous. The following lemma is then used.

Lemma 9 *Suppose that* $a \lhd_{\mathcal{X}} U_i \cup F_i^{-1}V$, *for every* $i \in I$. *Then* $a \lhd_{\mathcal{X}} W \cup F^{-1}V$.

Proof By localisation, we have, since U_i is down-closed,

$$a \lhd_{\mathcal{X}} \bigcap_{i\in I}(U_i \cup F_i^{-1}V).$$

Let x be an element in the right hand side. Using that I is finite, there are only two cases to consider:

Case 1: If $x \in U_i$ for all i, we have $x \in \bigcap_{j\in I} U_j = W$, and, trivially, $x \lhd_{\mathcal{X}} W \cup F^{-1}V$.

Case 2: If $x \in F_i^{-1}V$, for some i, then V is inhabited and thus $U_j \subseteq F_j^{-1}V$ for any j, which implies that $x \in \bigcap_{j\in I} F_j^{-1}V$. Hence there are $b_j \in V$, $j \in J$, so that $x\ F_j\ b_j$ for each $j \in J$. Fix $j \in J$. Using (8) it follows from $x\ F_j\ b_j$, that for any k, $x \lhd_{\mathcal{X}} U_j \cup F_k^{-1}b_j$. Thus by localization, and the definition of F,

$$x \lhd_{\mathcal{X}} U_j \cup \bigcap_{k\in I} F_k^{-1}b_j = U_j \cup F^{-1}b_j \subseteq U_j \cup F^{-1}V$$

Again by localisation, we obtain $x \lhd_{\mathcal{X}} (\bigcap_{j\in I} U_j) \cup F^{-1}V = W \cup F^{-1}V$ as desired. □

Using this lemma the conditions (A1) – (A4) for continuity are now straightforward to check. This finishes the proof of the glueing theorem. □

The theorem now gives the following corollary which appeared in [18].

Corollary 10 *Let $\mathcal{Y}$ be a formal topology and suppose that $U_1, U_2 \subseteq Y$ are saturated subsets. Write $V = U_1 \cup U_2$ and $W = U_1 \cap U_2$. Then the following diagram is a pushout*

$$\begin{tikzcd} \mathcal{Y}\dot{-}V \arrow[r, "E_{V,U_2}"] \arrow[d, "E_{V,U_1}"'] & \mathcal{Y}\dot{-}U_2 \arrow[d, "E_{U_2,W}"] \\ \mathcal{Y}\dot{-}U_1 \arrow[r, "E_{U_1,W}"'] & \mathcal{Y}\dot{-}W \end{tikzcd} \tag{12}$$

Proof By (4) it follows that the diagram commutes. The pushout property is a direct consequence of Theorem 8 for $I = \{1, 2\}$. □

We use this result to prove HJP for formal topology in Section 4.

3.3 Formal Reals

The basic neighbourhoods of the formal reals $\mathcal{R}$ are $R = \{(a, b) \in \mathbb{Q}^2 : a < b\}$ given the inclusion order (as intervals), denoted by $\leqslant_{\mathcal{R}}$. The cover $\lhd_{\mathcal{R}}$ is generated by

(G1) $(a, b) \lhd \{(a', b') : a < a' < b' < b\}$ for all $a < b$,
(G2) $(a, b) \lhd \{(a, c), (d, b)\}$ for all $a < d < c < b$.

This means that $\lhd_{\mathcal{R}}$ is the smallest covering relation satisfying (G1) and (G2). The set of points $\mathrm{Pt}(\mathcal{R})$ form a structure isomorphic to the Cauchy reals $\mathbb{R}$. The order relation of points is given by $\alpha < \beta$ iff $b < c$ for some $(a, b) \in \alpha$ and $(c, d) \in \beta$. Define $\alpha \leqslant \beta$ iff $\neg(\beta < \alpha)$. The latter is equivalent to: $b \leqslant c$ for all $(a, b) \in \alpha$ and $(c, d) \in \beta$.

For a rational $q \in \mathbb{Q}$ define the corresponding real by $\check{q} = \{(a, b) \in R : a < q < b\}$. We write $\check{q}$ as q when no confusion can arise. It is easily seen that $(a, b)^* \subseteq (c, d)^*$ implies $(a, b) \leqslant (c, d)$. By Lemma 6 its follows that $\mathcal{R}$ is point-wise order conservative.

The following may be regarded as a point-free version of the trichotomy principle for real numbers.

Theorem 11 *For any point* β *of the formal reals* $\mathcal{R}$ *let*

$$T_\beta = \{(a,b) \in R : b < \beta \text{ or } (a,b) \in \beta \text{ or } \beta < a\}.$$

Then for any $U \subseteq R$ *we have*

$$U \sim_{\mathcal{R}} U_{\leqslant} \cap T_\beta.$$

Proof The covering $U_{\leqslant} \cap T_\beta \lhd_{\mathcal{R}} U$ is clear by the axioms (Tra) and (Ext). To prove the converse covering, suppose that $(a,b) \in U$. Then it suffices by axiom (G1) to show $(c,d) \lhd_{\mathcal{R}} U_{\leqslant} \cap T_\beta$ for any $a < c < d < b$. For any such c, d we have $\beta < c$ or $a < \beta$, by the co-transitivity principle for real numbers. In the former case, $(c,d) \in U_{\leqslant} \cap T_\beta$, so suppose $a < \beta$. Similarly comparing d, b to β we get $d < \beta$, in which case $(c,d) \in U_{\leqslant} \cap T_\beta$, or we get $\beta < b$. In the latter case we have $(a,b) \in U_{\leqslant} \cap T_\beta$, so in particular $(c,d) \lhd U_{\leqslant} \cap T_\beta$. □

Next we define the closed interval $[\alpha, \beta]$, for $\alpha \leqslant \beta$, as the formal topology $\mathcal{R} \dot{-} U_{\alpha,\beta}$ where

$$U_{\alpha,\beta} = \{(a,b) : b < \alpha \text{ or } \beta < a\}.$$

Then $[\alpha, \alpha]$ is a terminal object in the category, but the proof of this is not as immediate as in point-set topology:

Lemma 12 *Let* $\mathcal{X}$ *be a formal topology. Let* $\beta \in \mathrm{Pt}(\mathcal{R})$. *Then* $\hat{\beta}$ *is the unique continuous map* $\mathcal{X} \longrightarrow [\beta, \beta]$.

Proof The proof that $\hat{\beta}$ is a continuous map $\mathcal{X} \longrightarrow [\beta, \beta]$ is left to the reader. Suppose now that $F \colon \mathcal{X} \longrightarrow [\beta, \beta]$ is another continuous map. For any $I \in \beta$, we have $R \lhd_{[\beta,\beta]} I$, and hence $F^{-1}R \lhd_{[\beta,\beta]} F^{-1}I$. Thus for any $u \in R$, we have, using (A3) for F, $u \lhd_{[\beta,\beta]} F^{-1}I$, i.e. $u\, F\, I$. It has been shown that for any I,

$$\{u \in R : I \in \beta\} \subseteq \{u \in R : u\, F\, I\} = F^{-1}I. \tag{13}$$

Suppose $x\, \hat{\beta}\, I$, i.e. $x \lhd_{\mathcal{X}} \{u \in R : I \in \beta\}$. Hence $x\, F\, I$, by (13). This proves $\hat{\beta} \subseteq F$.

To prove $F \subseteq \hat{\beta}$, suppose that $x\, F\, I$. By Theorem 11,

$$I \sim_{\mathcal{R}} I_{\leqslant} \cap T_\beta.$$

Thus $x \lhd_{\mathcal{X}} F^{-1}(I_{\leqslant} \cap T_\beta)$. Take any $u \in F^{-1}(I_{\leqslant} \cap T_\beta)$. It suffices to prove $u\, \hat{\beta}\, I$. For some $J = (a,b) \leqslant I$ we have $J \in T_\beta$ and $u\, F\, J$. There are three cases: (i) $b < \beta$, (ii) $\beta < a$ and (iii) $a < \beta < b$. For (i) we get $J \lhd_{[\beta,\beta]} \emptyset$, so

$$u \lhd_{\mathcal{X}} F^{-1}J \lhd_{\mathcal{X}} F^{-1}\emptyset = \emptyset = \{v \in R : J \in \beta\},$$

i.e. $u\,\hat{\beta}\,J$. Case (ii) is symmetric and yields $u\,\hat{\beta}\,J$ as well. In case (iii), $J \in \beta$, so $u\,\hat{\beta}\,J$ is immediate. Thus in all cases $u\,\hat{\beta}\,J$. This shows

$$F^{-1}(I_{\leqslant} \cap T_{\beta}) \subseteq \hat{\beta}^{-1}(I_{\leqslant} \cap T_{\beta}),$$

and therefore by transitivity $x \lhd_{\mathcal{X}} \hat{\beta}^{-1}(I_{\leqslant} \cap T_{\beta})$. Again using $I \sim_{\mathcal{R}} I_{\leqslant} \cap T_{\beta}$, we get $x \lhd_{\mathcal{X}} \hat{\beta}^{-1}(I)$, i.e. $x\,\hat{\beta}\,I$. This shows $F \subseteq \hat{\beta}$. □

3.4 Product Topologies

We recall the construction of the product of two formal topologies $\mathcal{X}_1 = (X_1, \leqslant_1, \lhd_1)$ and $\mathcal{X}_2 = (X_2, \leqslant_2, \lhd_2)$. The product is $\mathcal{X} = (X_1 \times X_2, \leqslant, \lhd)$ where $(x_1, x_2) \leqslant (y_1, y_2)$ iff $x_1 \leqslant_1 y_1$ and $x_2 \leqslant y_2$, and where $\lhd$ is the least cover relation so that

(PC1) $x_1 \lhd_1 U$ implies $(x_1, x_2) \lhd U \times \{x_2\}$
(PC2) $x_2 \lhd_2 V$ implies $(x_1, x_2) \lhd \{x_1\} \times V$.

The projections $P_1 : \mathcal{X} \longrightarrow \mathcal{X}_1$ and $P_2 : \mathcal{X} \longrightarrow \mathcal{X}_2$ are given by

$$(x_1, x_2)\, P_1\, u \Leftrightarrow (x_1, x_2) \lhd \{u\} \times X_2$$
$$(x_1, x_2)\, P_2\, v \Leftrightarrow (x_1, x_2) \lhd X_1 \times \{v\}.$$

If $F: \mathcal{Z} \longrightarrow \mathcal{X}_1$ and $G:\mathcal{Z} \longrightarrow \mathcal{X}_2$ are continuous, then $\langle F, G\rangle : \mathcal{Z} \longrightarrow \mathcal{X}$ given by

$$z\, \langle F, G\rangle\, (x_1, x_2) \Longleftrightarrow_{\text{def}} z\, F\, x_1 \text{ and } z\, G\, x_2$$

is the unique continuous map $\mathcal{Z} \longrightarrow \mathcal{X}$ such that $P_1 \circ \langle F, G\rangle = F$ and $P_2 \circ \langle F, G\rangle = G$. For $H_1 : \mathcal{Z}_1 \longrightarrow \mathcal{X}_1$ and $H_2 : \mathcal{Z}_2 \longrightarrow \mathcal{X}_2$ we write as usual $H_1 \times H_2$ for $\langle H_1 \circ P_1, H_2 \circ P_2\rangle$. The following lemma is easily proved by induction on the covers of the products.

Lemma 13 *Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be formal topologies. Suppose that $(x_1, x_2) \lhd_{\mathcal{X}_1 \times \mathcal{X}_2} W$.*

(i) If $x_2 \in \beta$ and $\beta \in \mathrm{Pt}(\mathcal{X}_2)$, then $x_1 \lhd_{\mathcal{X}_1} \{y_1 \in X_1 : (\exists y_2 \in \beta)(y_1, y_2) \in W\}$.
(ii) If $x_1 \in \alpha$ and $\alpha \in \mathrm{Pt}(\mathcal{X}_1)$, then $x_2 \lhd_{\mathcal{X}_2} \{y_2 \in X_2 : (\exists y_1 \in \alpha)(y_1, y_2) \in W\}$.

In particular, if $(x_1, x_2)\, P_1\, y$ and x_2 belongs to some point of $\mathcal{X}_2$, then $x_1 \lhd_{\mathcal{X}_1} y$. On the other hand, if $(x_1, x_2)\, P_2\, y$ and x_1 belongs to some point of $\mathcal{X}_1$, then $x_2 \lhd_{\mathcal{X}_2} y$.

Corollary 14 *Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be order conservative formal topologies, where every neighbourhood contains a point. Then the projection $P_k : \mathcal{X}_1 \times \mathcal{X}_2 \longrightarrow \mathcal{X}_k$ satisfies*

$$(x_1, x_2)\, P_k\, y \Longleftrightarrow x_k \leqslant y.$$

Using Lemma 13 it is straightforward to prove:

Proposition 15 *Let* $\mathcal{X}_1$ *and* $\mathcal{X}_2$. *Then* $\mathrm{Pt}(\mathcal{X}_1 \times \mathcal{X}_2)$ *and* $\mathrm{Pt}(\mathcal{X}_1) \times \mathrm{Pt}(\mathcal{X}_2)$ *are homeomorphic via* $\gamma \mapsto (\gamma_{(1)}, \gamma_{(2)})$ *and* $(\alpha, \beta) \mapsto \langle \alpha, \beta \rangle$. *Here* $\gamma_{(1)} = \{a : (\exists b)\,(a, b) \in \gamma\}$, $\gamma_{(2)} = \{b : (\exists a)\,(a, b) \in \gamma\}$ *and* $\langle \alpha, \beta \rangle = \{(a, b) : a \in \alpha, b \in \beta\}$.

On the other hand, it is well-known in classical locale theory that products are not preserved by the left adjoint Ω to Pt, unless at least one factor is locally compact [13, p. 61].

To calculate subspaces of a product space the following result is useful.

Lemma 16 *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be formal topologies.*

(a) *If* V *is a subset of* Y, *then the covers of the spaces* $\mathcal{X} \times (\mathcal{Y} \dot{-} V)$ *and* $\mathcal{X} \times \mathcal{Y} \dot{-} X \times V$, *are the same:*

$$w \lhd_{\mathcal{X} \times (\mathcal{Y} \dot{-} V)} W \Longleftrightarrow w \lhd_{\mathcal{X} \times \mathcal{Y} \dot{-} X \times V} W.$$

(b) *For* $U \subseteq V \subseteq Y$ *the maps*

$$1_{\mathcal{X}} \times E_{V,U} : \mathcal{X} \times (\mathcal{Y} \dot{-} V) \longrightarrow \mathcal{X} \times (\mathcal{Y} \dot{-} U)$$

and

$$E_{X \times V, X \times U} : \mathcal{X} \times \mathcal{Y} \dot{-} X \times V \longrightarrow \mathcal{X} \times \mathcal{Y} \dot{-} X \times U$$

are identical.

Proof Part (a): $(\Rightarrow)$ is proved by induction on covers. It suffices to check the implication for generators of the product, which is straightforward.

$(\Leftarrow)$ is proved by first establishing the following implication by induction

$$w \lhd_{\mathcal{X} \times \mathcal{Y}} Z \Longrightarrow w \lhd_{\mathcal{X} \times (\mathcal{Y} \dot{-} V)} Z.$$

Using this, one gets from $w \lhd_{\mathcal{X} \times \mathcal{Y} \dot{-} X \times V} W$ that $w \lhd_{\mathcal{X} \times (\mathcal{Y} \dot{-} V)} (X \times V) \cup W$. But since $V \lhd_{\mathcal{Y} \dot{-} V} \emptyset$, we get by (PC2) in fact $w \lhd_{\mathcal{X} \times (\mathcal{Y} \dot{-} V)} W$.

Part (b): We have by the definitions of the two maps to be compared

$$(x, y)\, E_{X \times V, X \times U}\, (u, v) \Longleftrightarrow (x, y) \lhd_{\mathcal{X} \times \mathcal{Y}} X \times V \cup \{(u, v)\} \tag{14}$$

and

$$(x, y)\, 1_{\mathcal{X}} \times E_{V,U}\, (u, v) \Longleftrightarrow (x, y)\, P_1\, u \text{ and } (x, y)\, (E_{V,U} \circ P_2)\, v \tag{15}$$

Using Part (a) we get

$$(x, y)\, P_1\, u \Longleftrightarrow (x, y) \lhd_{\mathcal{X} \times \mathcal{Y}} X \times V \cup \{u\} \times Y. \tag{16}$$

Furthermore, a calculation gives

$$(x, y)\,(E_{V,U} \circ P_2)\,v \Longleftrightarrow (x, y) \lhd_{\mathcal{X}\times\mathcal{Y}} X \times V \cup X \times \{v\}. \tag{17}$$

By applying localisation to the right hand sides of (16) and (17) it follows that the right hand sides of (14) and (15) are equivalent. □

4 HJP for Formal Topologies

Given the preparations of the previous section we can now quite straightforwardly prove the main result.

Theorem 17 *Let $\mathcal{X}$ be a formal topology. For $\alpha \leqslant \beta \leqslant \gamma$ in* Pt($\mathcal{R}$)*, the diagram*

$$\begin{tikzcd}
\mathcal{X} \arrow[r, "\langle 1_{\mathcal{X}}, \hat{\beta}\rangle"] \arrow[d, "\langle 1_{\mathcal{X}}, \hat{\beta}\rangle"'] & \mathcal{X} \times [\beta, \gamma] \arrow[d, "1_{\mathcal{X}} \times E_2"] \\
\mathcal{X} \times [\alpha, \beta] \arrow[r, "1_{\mathcal{X}} \times E_1"'] & \mathcal{X} \times [\alpha, \gamma]
\end{tikzcd} \tag{18}$$

is a pushout diagram. Here E_1 and E_2 are the obvious embeddings of subspaces.

Proof We employ Corollary 10 with $\mathcal{Y} = \mathcal{X} \times \mathcal{R}$ as the formal topology, and with $U_1 = (X \times U_{\alpha,\beta})_{\lhd}$ and $U_2 = (X \times U_{\beta,\gamma})_{\lhd}$ as saturated subsets, writing $\lhd$ for $\lhd_{\mathcal{Y}}$. This gives diagram (12), where $V = U_1 \cup U_2$ and $W = U_1 \cap U_2$. We have

$$X \times U_{\beta,\beta} = X \times U_{\alpha,\beta} \cup X \times U_{\beta,\gamma}.$$

It follows easily that

$$V_{\lhd} = (X \times U_{\beta,\beta})_{\lhd} \tag{19}$$

Furthermore, we have

$$X \times U_{\alpha,\gamma} = X \times U_{\alpha,\beta} \cap X \times U_{\beta,\gamma}.$$

From which it follows that

$$W = (X \times U_{\alpha,\gamma})_{\lhd} \tag{20}$$

By (2) and (19) that

$$\mathcal{Y}\dot{-}V = \mathcal{Y}\dot{-}V_{\lhd} = \mathcal{Y}\dot{-}(X \times U_{\beta,\beta})_{\lhd} = \mathcal{Y}\dot{-}X \times U_{\beta,\beta}.$$

Then using Lemma 16.(a) the right hand side is seen to equal $\mathcal{X} \times (\mathcal{R}\dot{-}U_{\beta,\beta})$. Thus

$$\mathcal{Y}\dot{-}V = \mathcal{X} \times [\beta, \beta] \tag{21}$$

Similarly, using Lemma 16.(a) and (2) we get

$$\mathcal{Y}\dot{-}U_1 = \mathcal{X} \times [\alpha, \beta] \qquad \mathcal{Y}\dot{-}U_2 = \mathcal{X} \times [\beta, \gamma] \tag{22}$$

As for the final corner of (12), we get by (2), (20) and Lemma 16.(a)

$$\mathcal{Y}\dot{-}W = \mathcal{X} \times [\alpha, \gamma] \tag{23}$$

We now use (5) to rewrite the maps of the diagram. From Lemma 16.(b) and (19) follows that

$$E_{V,U_1} = E_{(X\times U_{\beta,\beta})_{\lhd},(X\times U_{\beta,\gamma})_{\lhd}} = E_{X\times U_{\beta,\beta},X\times U_{\beta,\gamma}} = 1_{\mathcal{X}} \times E_{U_{\beta,\beta},U_{\beta,\gamma}}. \tag{24}$$

Symmetrically,

$$E_{V,U_2} = 1_{\mathcal{X}} \times E_{U_{\beta,\beta},U_{\alpha,\beta}}. \tag{25}$$

Similarly, now using (20), we get

$$E_{U_1,W} = 1_{\mathcal{X}} \times E_{U_{\alpha,\beta},U_{\alpha,\gamma}} \qquad E_{U_2,W} = 1_{\mathcal{X}} \times E_{U_{\beta,\gamma},U_{\alpha,\gamma}}. \tag{26}$$

Abbreviating $E_1 = E_{U_{\alpha,\beta},U_{\alpha,\gamma}}$, $E_2 = E_{U_{\beta,\gamma},U_{\alpha,\gamma}}$, $E^1 = E_{U_{\beta,\beta},U_{\alpha,\beta}}$ and $E^2 = E_{U_{\beta,\beta},U_{\beta,\gamma}}$, we conclude from equations (21, 22, 23, 24, 25, 26) above that the following is then a pushout diagram:

$$\begin{tikzcd}
\mathcal{X} \times [\beta,\beta] \arrow[r, "1_{\mathcal{X}}\times E^2"] \arrow[d, "1_{\mathcal{X}}\times E^1"'] & \mathcal{X} \times [\beta,\gamma] \arrow[d, "1_{\mathcal{X}}\times E_2"] \\
\mathcal{X} \times [\alpha,\beta] \arrow[r, "1_{\mathcal{X}}\times E_1"'] & \mathcal{X} \times [\alpha,\gamma]
\end{tikzcd} \tag{27}$$

To obtain the pushout diagram (18) from (27) it suffices to show that $\langle 1_{\mathcal{X}}, \hat{\beta}\rangle : \mathcal{X} \longrightarrow \mathcal{X} \times [\beta, \beta]$ is an isomorphism and that

$$(1_{\mathcal{X}} \times E^1) \circ \langle 1_{\mathcal{X}}, \hat{\beta}\rangle = \langle 1_{\mathcal{X}}, \hat{\beta}\rangle \qquad (1_{\mathcal{X}} \times E^2) \circ \langle 1_{\mathcal{X}}, \hat{\beta}\rangle = \langle 1_{\mathcal{X}}, \hat{\beta}\rangle. \tag{28}$$

The equations (28) are straightforward to check. As for the isomorphism note that $P_1 \circ \langle 1_{\mathcal{X}}, \hat{\beta} \rangle = 1_{\mathcal{X}}$. Moreover, $\langle 1_{\mathcal{X}}, \hat{\beta} \rangle \circ P_1 = \langle 1_{\mathcal{X}} \circ P_1, \hat{\beta} \circ P_1 \rangle = \langle P_1, \hat{\beta} \rangle$. But $P_2 = \hat{\beta} : \mathcal{X} \times [\beta, \beta] \longrightarrow [\beta, \beta]$, by Lemma 12, so $\langle P_1, \hat{\beta} \rangle = \langle P_1, P_2 \rangle = 1_{\mathcal{X} \times [\beta,\beta]}$. $\square$

Remark 18 The special case of Theorem 17 when $\mathcal{X} = 1$ gives the PJP property for formal topologies. This property could as well be obtained more easily from Corollary 10, directly. The latter yields a simple alternative proof of Theorem 17, under the restrictive assumption that $\mathcal{X}$ is locally compact. It is known that then the exponentiation functor $(-)^x$ exists; see Maietti [24]. Hence by adjointness $\mathcal{X} \times (-)$ is functor which preserves colimits and, in particular, pushouts. Applying the functor to an arbitary PJP diagram then gives the result.

Acknowledgement The author is grateful for remarks by the referee that helped improve the presentation.

References

1. M.A. Armstrong. *Basic Topology.* Springer, 1985.
2. B. Banaschewski, T. Coquand and G. Sambin (eds.). Papers presented at the Second Workshop Formal Topology, Venice, April 2–4, 2002. Special issue of *Ann. Pure Appl. Logic* 137, 2006.
3. B. Banaschewski and C.J. Mulvey. A constructive proof of the Stone-Weierstrass theorem. *J. Pure Appl. Algebra* 116, 25–40, 1997.
4. M. Beeson. *Foundations of Constructive Mathematics.* Springer, 1985.
5. E. Bishop and D.S. Bridges. *Constructive Analysis.* Springer, 1985.
6. T. Coquand. An intuitionistic proof of Tychonoff's theorem. *J. Symbolic Logic* 57, 28–32, 1992.
7. L. Crosilla and P. Schuster (eds.). *From Sets and Types to Topology and Analysis: Towards Practicable Foundations of Constructive Mathematics.* Oxford Logic Guides, Oxford University Press, 2005.
8. G. Curi. *Geometry of Observations: Some Contributions to (Constructive) Point-Free Topology.* PhD Thesis, Siena, 2004.
9. H. Freudenthal. Zum intuitionistischen Raumbegriff. *Compositio Math.* 4, 82–111, 1936.
10. H. Freudenthal and A. Heyting. The Life of L.E.J. Brouwer. In: H. Freudenthal (ed.), *L.E.J. Brouwer, Collected Works, Vol. 2.* North-Holland, 1976.
11. M.J. Greenberg. *Lectures on Algebraic Topology.* Benjamin, New York, 1967.
12. W. He. Homotopy theory for locales. (Chinese). *Acta Math. Sinica* 46, (5), 951–960. (ISSN 0583-1431), 2003.
13. P.T. Johnstone. *Stone Spaces.* Cambridge University Press, 1982.
14. P.T. Johnstone. The point of pointless topology. *Bull. Amer. Math. Soc.* 8, 41–53, 1983.
15. A. Joyal and M. Tierney. An extension of the Galois theory of Grothendieck. *Memoirs Amer. Math. Soc.* 309, 1984.
16. J.F. Kennison. What is the fundamental group? *J. Pure Appl. Algebra* 59, 187–200, 1989.
17. P. Martin-Löf. *Notes on Constructive Mathematics.* Almqvist and Wiksell, 1970.
18. E. Palmgren. Predicativity problems in point-free topology. In: V. Stoltenberg-Hansen and J. Vänänen (eds.), *Proceedings of the Annual European Summer Meeting of the Association for Symbolic Logic, held in Helsinki, Finland, August* 14–20, 2003, Lecture Notes in Logic 24, AK Peters, 2006.
19. E. Palmgren. Continuity on the real line and in formal spaces. In: *From Sets and Types to Topology and Analysis: Towards practicable Foundations of Constructive Mathematics*. Oxford Logic Guides, Oxford University Press, 2005.

20. E. Palmgren. Regular universes and formal spaces, *Ann. Pure Appl. Logic* 137, 2006.
21. G. Sambin. Intuitionistic formal spaces — a first communication. In: D. Skordev (ed.), *Mathematical Logic and its Applications.* Plenum Press, pp. 187–204, 1987.
22. A.S. Troelstra. *Intuitionistic General Topology.* PhD Thesis, Amsterdam, 1966.
23. F. Waaldijk. *Modern Intuitionistic Topology.* PhD Thesis, Nijmegen, 1998.
24. M.E. Maietti. Predicative exponentiation of locally compact formal topologies over inductively generated ones. In: L. Crosilla and P. Schuster (eds.), *From Sets and Types to Topology and Analysis: Towards Practicable Foundations of Constructive Mathematics*, Oxford Logic Guides, Oxford University Press 2005.

Program Extraction in Constructive Analysis

Helmut Schwichtenberg

Abstract We sketch a development of constructive analysis in Bishop's style, with special emphasis on low type-level witnesses (using separability of the reals). The goal is to set up things in such a way that realistically executable programs can be extracted from proofs. This is carried out for (1) the Intermediate Value Theorem and (2) the existence of a continuous inverse to a monotonically increasing continuous function. Using the Minlog proof assistant, the proofs leading to the Intermediate Value Theorem are formalized and realizing terms extracted. It turns out that evaluating these terms is a reasonably fast algorithm to compute, say, approximations of $\sqrt{2}$.

1 Introduction

We are interested in *exact real numbers*, as opposed to floating point numbers. The final goal is to develop the basics of real analysis in such a way that from a proof of an existence formula one can extract a program. For instance, from a proof of the intermediate value theorem we want to extract a program that, given an arbitrary error bound 2^{-k}, computes a rational a where the given function is zero up to the error bound.

Why should we be interested in logic in a study of constructive analysis? There are at least two reasons.

- Obviously we need to be aware of the difference of the classical and the constructive existential quantifier, and try to prove the stronger statements involving the latter whenever possible. Then one is forced to give "constructive" proofs, whose algorithmic content can be "seen" and then used as a basis to formulate a program for computing the solution. This was the point of view in Bishop's classic textbook [4], and more explicitly carried through in Andersson's Master's thesis [1] (based on Palmgren's [9]), with Mathematica as the target programming language.

H. Schwichtenberg (✉)
Mathematisches Institut der Universität Münche, München, Germany
e-mail: schwicht@math.lmu.de

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_13, © Springer Science+Business Media B.V. 2009

- However, one can go one step further and automatize the step from the (formalized) constructive proof to the corresponding program. This can be done by means of the so-called realizability interpretation introduced by Kreisel [7]. The desire to have "mathematics as a numerical language" in this sense was clearly expressed by Bishop in his article [5] (with just that title). There are now many implementations of these ideas, for instance Coq, Isabelle/HOL, Agda, Nuprl, Minlog (cf. Wiedijk's "The Seventeen Provers of the World", to appear as Springer LNAI).

What are the requirements on a constructive logic that should guide us in our design?

- It should be as close as possible to the mathematical arguments we want to use. Variables should carry (functional) types, with free algebras (e.g., natural numbers) as base types. Over these, inductive definitions and the corresponding introduction and elimination axioms should be allowed.
- The constants of the language should denote computable functionals in the Scott-Ershov sense, and hence the higher-order quantifiers should range over their (mathematically correct) domain, the partial continuous functionals.
- The language of the logic should be strong (in the sense of being expressive), but the existence axioms used should be weak.
- Type parameters (ML style) should be allowed, but quantification over types should be disallowed in order to keep the theory predicative. Similarly, predicate variables should be allowed as place-holders for formulas (or more precisely, comprehension terms, that is formulas with some variables abstracted). However, in comprehension terms quantification over predicate variables is not allowed, since this would form a glaring impredicativity: we then would define a predicate (by the comprehension term) with reference to the totality of all predicates, to which the one to be defined belongs.

The Minlog proof assistant (www.minlog-system.de, under development in Munich), has been designed to meet these demands. Some information on the logical theory it formalizes can be found in [11]. The present paper reports on experiences in formalizing proofs in constructive analysis and extracting programs from them, using the Minlog proof assistant.

Sections 2 and 3 develop some basic machinery of constructive analysis, in a form suitable for formalization. This includes proofs of the Intermediate Value Theorem in Section 3.4 and of the fact that every continuous function with a uniform lower bound on its slope has a continuous inverse, in Section 3.5. In Section 4 we report on a formalization of the material in Sections 2 and 3, which includes program extraction. After providing in Section 4.1 some general proof-theoretical background on program extraction and on necessary optimizations, in Section 4.2 we display the extracted terms for the proofs leading to the Intermediate Value Theorem, and discuss what they intuitively mean. We then report on experiments using these terms to compute approximations of $\sqrt{2}$.

Related work has recently been done in the "Constructive Coq Repository (C-CoRN)" project at Nijmegen (Barendregt, Geuvers, Wiedijk, Cruz-Filipe [6]). It grew out of the FTA project, where Kneser's constructive proof of the fundamental theorem of algebra was to be formalized. However, program extraction from this proof turned out to be problematic, for a number of reasons, among them:

- Strong extensionality was required, in the form $\forall_{x,y}.f(x)\#f(y) \to x\#y$. However, the missing witness in this formula turned out to be harmful for program extraction.
- The **Set**, **Prop** distinction in Coq was found to be insufficient. In his thesis [6], Cruz-Filipe was forced to introduce a certain variant he called "**CProp**".

Compared with the literature, the novel aspect of the present work is the development of elementary constructive analysis in such a way that witnesses have as low a type level as possible. For example, a continuous function on the reals is determined by its values on the rationals, and hence can be represented by a type-one (rather than type-two) object. This clearly is important for the complexity of the extracted programs.

The subject of the present volume is "Logicism, Intuitionism, and Formalism – What has become of them?" This paper can be seen as providing first steps in an attempt to develop "Mathematics as a numerical language", as advocated by Bishop [5]. Clearly this requires a constructive view of mathematics as advocated by Brouwer, but in the liberal sense of Bishop and Troelstra. But of course, any such attempt would be impossible without the pioneering ideas and concepts from all these foundational schools.

2 Real Numbers

2.1 Reals, Equality of Reals

We shall view a real as a Cauchy sequence of rationals with a separately given modulus.

Definition A real number x is a pair $((a_n)_{n\in\mathbb{N}}, M)$ with $a_n \in \mathbb{Q}$ and $M\colon \mathbb{N} \to \mathbb{N}$ such that $(a_n)_n$ is a Cauchy sequence with modulus M, that is

$$|a_n - a_m| \leqslant 2^{-k} \quad \text{for } n, m \geqslant M(k)$$

and M is weakly increasing. M is called a Cauchy modulus of x.

We shall loosely speak of a real $(a_n)_n$ if the Cauchy modulus M is clear from the context or inessential. Every rational a is tacitly understood as the real represented by the constant sequence $a_n = a$ with the constant modulus $M(k) = 0$.

Definition Two reals $x := ((a_n)_n, M)$, $y := ((b_n)_n, N)$ are called *equivalent* (or *equal*), which is written $x = y$ if the context makes clear what is meant, if

$$|a_{M(k+1)} - b_{N(k+1)}| \leqslant 2^{-k} \quad \text{for all } k \in \mathbb{N}$$

We want to show that this is an equivalence relation. Reflexivity and symmetry are clear. For transitivity we use the following lemma:

Two reals $x := ((a_n)_n, M)$, $y := ((b_n)_n, N)$ are equal if and only if

$$\forall_k \exists^{\mathrm{cl}}_q \forall_{n \geqslant q} |a_n - b_n| \leqslant 2^{-k}.$$

Lemma *Equality between reals is transitive.*

Proof Let $(a_n)_n$, $(b_n)_n$, $(c_n)_n$ be the Cauchy sequences for x, y, z. Assume $x = y$, $y = z$ and pick p, q according to the lemma above. Then $|a_n - c_n| \leqslant |a_n - b_n| + |b_n - c_n| \leqslant 2^{-k-1} + 2^{-k-1}$ for $n \geqslant p, q$.

2.2 The Archimedean Axiom

For every function on the reals we certainly want compatibility with equality. This however is not always the case; here is an important example.

Lemma (RealBound) *For every real $x := ((a_n)_n, M)$ we can find an upper bound 2^{k_x} on the elements of the Cauchy sequence: $|a_n| \leqslant 2^{k_x}$ for all n.* □

Proof Let k_x be such that $\max\{\,|a_n| \mid n \leqslant M(0)\,\} + 1 \leqslant 2^{k_x}$. Hence $|a_n| \leqslant 2^{k_x}$ for all n.

Clearly this assignment of k_x to x is not compatible with equality.

2.3 Nonnegative and Positive Reals

A real $x := ((a_n)_n, M)$ is called *nonnegative* (written $x \in \mathrm{d}R^{0+}$) if

$$-2^{-k} \leqslant a_{M(k)} \quad \text{for all } k \in \mathbb{N}.$$

It is *k-positive* (written $x \in_k \mathbb{R}^+$, or $x \in \mathbb{R}^+$ if k is not needed) if

$$2^{-k} \leqslant a_{M(k+1)}.$$

We want to show that both properties are compatible with equality. First we prove a useful characterization of nonnegative reals.

Lemma *A real $x := ((a_n)_n, M)$ is nonnegative if and only if*

$$\forall_k \exists_p \forall_{n \geqslant p} - 2^{-k} \leqslant a_n.$$

Lemma *If $x \in \mathbb{R}^{0+}$ and $x = y$, then also $y \in \mathbb{R}^{0+}$.*

Proof Let $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$. Assume $x \in \mathbb{R}^{0+}$ and $x = y$, and let k be given. Pick p according to the lemma above and q according to the characterization of equality of reals in Section 2.1 (both for $k+1$). Then for $n \geqslant p, q$

$$-2^{-k} \leqslant -2^{-k-1} + a_n \leqslant (b_n - a_n) + a_n.$$

Hence $y \in \mathbb{R}^{0+}$ by definition. □

We now show compatibility of positivity with equality. Again we need a lemma:

Lemma *A real $x := ((a_n)_n, M)$ is k-positive if and only if*

$$\exists_{l,p} \forall_{n \geqslant p} \, 2^{-l} \leqslant a_n.$$

For $\forall_{n \geqslant p} \, 2^{-l} \leqslant a_n$ write $x \in_{l,p} \mathbb{R}^+$ or $0 <_{l,p} x$.

Lemma *Positivity of reals is compatible with equality.*

Proof Assume $0 <_{k,p} x$ and $x = y$, so in particular we have a q such that $\forall_{n \geqslant q} \, |a_n - b_n| \leqslant 2^{-k-1}$. Then for $\max(p, q) \leqslant n$

$$2^{-k-1} = -2^{-k-1} + 2^k \leqslant (b_n - a_n) + a_n = b_n.$$

Hence $0 <_{k+1,\max(p,q)} y$. □

2.4 Arithmetical Functions

Given real numbers $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$, we define each z from the list $x + y$, $-x$, $|x|$, $x \cdot y$, and $\frac{1}{x}$ (the latter only provided that $|x| \in_l \mathbb{R}^+$) as represented by the respective sequence (c_n) of rationals with modulus L:

z	c_n	$L(k)$
$x + y$	$a_n + b_n$	$\max\big(M(k+1), N(k+1)\big)$
$-x$	$-a_n$	$M(k)$
$\|x\|$	$\|a_n\|$	$M(k)$
$x \cdot y$	$a_n \cdot b_n$	$\max\big(M(k+1+k_{\|y\|}), N(k+1+k_{\|x\|})\big)$
$\frac{1}{x}$ for $\|x\| \in_l \mathbb{R}^+$	$\begin{cases} \frac{1}{a_n} & \text{if } a_n \neq 0 \\ 0 & \text{if } a_n = 0 \end{cases}$	$M(2(l+1)+k)$,

where 2^{k_x} is the upper bound provided by Section 2.2.

Lemma *For reals x, y also $x + y$, $-x$, $|x|$, $x \cdot y$ and (provided that $|x| \in_l \mathbb{R}^+$) also $1/x$ are reals.*

Lemma *The functions $x + y$, $-x$, $|x|$, $x \cdot y$ and (provided that $|x| \in_l \mathbb{R}^+$) also $1/x$ are compatible with equality.*

Lemma *For reals x, y, z*

$$x + (y + z) = (x + y) + z \qquad x \cdot (y \cdot z) = (x \cdot y) \cdot z$$
$$x + 0 = x \qquad x \cdot 1 = x$$
$$x + (-x) = 0 \qquad 0 < |x| \to x \cdot \frac{1}{x} = 1$$
$$x + y = y + x \qquad x \cdot y = y \cdot x$$
$$x \cdot (y + z) = x \cdot y + x \cdot z$$

Lemma *For reals x, y from $x \cdot y = 1$ we can infer $0 < |x|$.*

Proof Pick k such that $|b_n| \leqslant 2^k$ for all n. Pick q such that $q \leqslant n$ implies $1/2 \leqslant a_n \cdot b_n$. Then for $q \leqslant n$, $1/2 \leqslant |a_n| \cdot 2^k$, and hence $2^{-k-1} \leqslant |a_n|$. □

Lemma *For reals x, y,*

(a) $x, y \in \mathbb{R}^{0+} \to x + y, x \cdot y \in \mathbb{R}^{0+}$,
(b) $x, y \in \mathbb{R}^{+} \to x + y, x \cdot y \in \mathbb{R}^{+}$,
(c) $x \in \mathbb{R}^{0+} \to -x \in \mathbb{R}^{0+} \to x = 0$.

Proof (a), (b). Routine. (c). Let k be given. Pick p such that $-2^{-k} \leqslant a_n$ and $-2^{-k} \leqslant -a_n$ for $n \geqslant p$. Then $|a_n| \leqslant 2^{-k}$.

2.5 Comparison of Reals

We write $x \leqslant y$ for $y - x \in \mathbb{R}^{0+}$ and $x < y$ for $y - x \in \mathbb{R}^+$. Unwinding the definitions yields that $x \leqslant y$ is to say that for every k, $a_{L(k)} \leqslant b_{L(k)} + 2^{-k}$ with $L(k) := \max(M(k), N(k))$, or equivalently (using the characterization of reals in Section 2.3) that for every k there exists p such that $a_n \leqslant b_n + 2^{-k}$ for all $n \geqslant p$. Furthermore, $x < y$ is a shorthand for the presence of k with $a_{L(k+1)} + 2^{-k} \leqslant b_{L(k+1)}$ with L the maximum of M and N, or equivalently (using the characterization of reals in Section 2.3) for the presence of k, q with $a_n + 2^{-k} \leqslant b_n$ for all $n \geqslant q$; we then write $x <_k y$ (or $x <_{k,q} y$) whenever we want to call these witnesses.

Lemma (RealApprox) $\forall_{x,k} \exists_a \; |a - x| \leqslant 2^{-k}$.

Proof Let $x = ((a_n), M)$. Given k, pick $a_{M(k)}$. We show $|a_{M(k)} - x| \leqslant 2^{-k}$, that is $|a_{M(k)} - a_{M(l)}| \leqslant 2^{-k} + 2^{-l}$ for every l. But this follows from $|a_{M(k)} - a_{M(l)}| \leqslant |a_{M(k)} - a_{M(k+l)}| + |a_{M(k+l)} - a_{M(l)}| \leqslant 2^{-k} + 2^{-l}$.

Lemma $0 \leqslant x$ *and* $0 <_k y$ *imply* $0 <_{k+1} x + y$.

Proof From $0 \leqslant x$ we have $\forall_l \exists_p \forall_{n \geqslant p}\ -2^{-l} \leqslant a_n$. From $0 <_k y$ we have some q such that $\forall_{n \geqslant q}\, 2^{-k} \leqslant b_n$. Pick p for $k+1$. Then $p, q \leqslant n$ implies $0 \leqslant a_n + 2^{-k-1}$ and $2^{-k-1} \leqslant b_n - 2^{-k-1}$, hence $2^{-k-1} \leqslant a_n + b_n$.

Lemma *For reals* x, y, z,

$$
\begin{array}{ll}
x \leqslant x & \\
x \leqslant y \to y \leqslant x \to x = y & x \not< x \\
x \leqslant y \to y \leqslant z \to x \leqslant z & x < y \to y < z \to x < z \\
x \leqslant y \to x + z \leqslant y + z & x < y \to x + z < y + z \\
x \leqslant y \to 0 \leqslant z \to x \cdot z \leqslant y \cdot z & x < y \to 0 < z \to x \cdot z < y \cdot z
\end{array}
$$

Proof From Section 2.4.

Lemma $x \leqslant y \to y <_k z \to x <_{k+1} z$.

Proof This follows from the next to last lemma.

As is to be expected in view of the existential and universal character of the predicates $<$ and $\leqslant$ on the reals, we have:

Lemma $x \leqslant y \leftrightarrow y \not< x$.

Proof $\to$. Assume $x \leqslant y$ and $y < x$. By the previous lemma we obtain $x < x$, a contradiction.

$\leftarrow$. It clearly suffices to show $0 \not< z \to z \leqslant 0$, for a real z given by $(c_n)_n$. Assume $0 \not< z$. We must show $\forall_k \exists_p \forall_{n \geqslant p}\, c_n \leqslant 2^{-k}$. Let k be given. By assumption $0 \not< z$, hence $\neg \exists_l\, 2^{-l} \leqslant c_{M(l+1)}$, hence $\forall_l\, c_{M(l+1)} < 2^{-l}$. For $l := k+1$ this implies $c_{M(k+2)} < 2^{-k-1}$, hence $c_n \leqslant c_{M(k+2)} + 2^{-k-2} < 2^{-k}$ for $M(k+2) \leqslant n$.

Constructively, we cannot compare two reals, but we can compare a real with a nontrivial interval ("Approximate Splitting Principle"):

Lemma (ApproxSplit) *Let* x, y, z *be given and assume* $x < y$. *Then either* $z \leqslant y$ *or* $x \leqslant z$.

Proof Let $x := ((a_n)_n, M)$, $y := ((b_n)_n, N)$, $z := ((c_n)_n, L)$. Assume $x <_k y$, that is (by definition) $1/2^k \leqslant b_p - a_p$ for $p := \max(M(k+2), N(k+2))$. Let $q := \max(p, L(k+2))$ and $d := (b_p - a_p)/4$.

Case $c_q \leqslant \frac{a_p + b_p}{2}$. We show $z \leqslant y$. It suffices to prove $c_n \leqslant b_n$ for $n \geqslant q$. To see this, observe

$$c_n \leqslant c_q + \frac{1}{2^{k+2}} \leqslant \frac{a_p + b_p}{2} + \frac{b_p - a_p}{4} = b_p - \frac{b_p - a_p}{4} \leqslant b_p - \frac{1}{2^{k+2}} \leqslant b_n.$$

Case $c_q \not\leqslant \frac{a_p + b_p}{2}$. We show $x \leqslant z$, via $a_n \leqslant c_n$ for $n \geqslant q$.

$$a_n \leqslant a_p + \frac{1}{2^{k+2}} \leqslant a_p + \frac{b_p - a_p}{4} \leqslant \frac{a_p + b_p}{2} - \frac{b_p - a_p}{4} \leqslant c_q - \frac{1}{2^{k+2}} \leqslant c_n.$$

This concludes the proof.

Notice that the boolean object determining whether $z \leqslant y$ or $x \leqslant z$ depends on the representation of x, y and z. In particular this assignment is not compatible with our equality relation.

One might think that the non-available comparison of two reals could be circumvented by using a maximum function. Indeed, such a function can easily be defined (component-wise), and it has the expected properties $x, y \leqslant \max(x, y)$ and $x, y \leqslant z \to \max(x, y) \leqslant z$. What is missing is the knowledge that $\max(x, y)$ equals one of its arguments, i.e., we do not have $\max(x, y) = x \vee \max(x, y) = y$.

3 Continuous Functions

We define continuous functions as type-1-objects, and prove their basic properties. This includes proofs of the Intermediate Value Theorem in Section 3.4 and of the fact that every continuous function with a uniform lower bound on its slope has a continuous inverse, in Section 3.5.

An inhabited, closed finite interval is called *compact*. We use I, J to denote compact intervals $[a, b]$ with rational end points $a < b$.

3.1 Continuous Functions as Type-1-Objects

Definition A *continuous function* $f : I \to \mathbb{R}$ is given by

(a) an approximating map $h_f : (I \cap \mathbb{Q}) \times \mathbb{N} \to \mathbb{Q}$ and a map $\alpha_f : \mathbb{N} \to \mathbb{N}$ such that $(h_f(a, n))_n$ is a Cauchy sequence with (uniform) modulus α_f;
(b) a modulus $\omega_f : \mathbb{N} \to \mathbb{N}$ of (uniform) continuity, which satisfies

$$|a - b| \leqslant 2^{-\omega_f(k)+1} \to |h_f(a, n) - h_f(b, n)| \leqslant 2^{-k} \quad \text{for } n \geqslant \alpha_f(k);$$

(c) a lower bound N_f and an upper bound M_f for all $h_f(a, n)$.

α_f and ω_f are required to be weakly increasing (that is, $n \leqslant m \to \alpha_f(n) \leqslant \alpha)f(m))$. A function real-valued function defined on an arbitrary interval is continuous if it is continuous on every compact subinterval of with rational end points.

Notice that a continuous function is given by objects of type level $\leqslant 1$ only. This is due to the fact that it suffices to define its values on rational numbers.

The lower and upper bound of the values of the approximating map have been included to ease the definition of composition of continuous functions; however, they also have an effect on computational efficiency.

An example is the *exponential function*. On $[c, d]$ $(c < d)$ it is given by

(a) the approximating map

$$h_{\exp}(a, n) := \sum_{k=0}^{n} \frac{a^k}{k!},$$

and a uniform Cauchy modulus $\alpha_{\exp}$, which can be computed from

$$\Big|\sum_{k=0}^{n} \frac{a^k}{k!} - \sum_{k=0}^{m} \frac{a^k}{k!}\Big| = \Big|\sum_{k=n+1}^{m} \frac{a^k}{k!}\Big| \leqslant \sum_{k=n+1}^{\infty} \frac{|a|^k}{k!} \leqslant 2\frac{|a|^{n+1}}{(n+1)!}$$

for $|a| \leqslant 1 + \frac{n}{2}$ and $n \leqslant m$;

(b) the modulus of uniform continuity, which can be obtained from

$$\begin{aligned}\Big|\sum_{k=0}^{n} \frac{a^k}{k!} - \sum_{k=0}^{n} \frac{b^k}{k!}\Big| &= \Big|\sum_{k=1}^{n} \frac{a^k - b^k}{k!}\Big| = |a-b| \sum_{k=1}^{n} \frac{1}{k!}\Big|\sum_{l=0}^{k-1} a^{k-1-l} b^l\Big| \\ &\leqslant |a-b| \sum_{k=1}^{n} \frac{kM^{k-1}}{k!} = |a-b| \sum_{k=0}^{n-1} \frac{M^k}{k!} < |a-b| \exp(M),\end{aligned}$$

where $M = \max(|c|, |d|)$;

(c) the lower bound $N_{\exp} := 0$ and upper bound

$$M_{\exp} := \sum_{k=0}^{L} \frac{d^k}{k!} + 2\frac{|d|^{L+1}}{(L+1)!} \quad \text{with } L := 2\lceil |d| \rceil;$$

these can easily be verified.

3.2 Application of a Continuous Function to a Real

Since the approximating map operates on rationals only, we need to define what it means to apply a continuous function in our sense to a real.

Definition *Application* of a continuous function $f : I \to \mathbb{R}$ (given by h_f, α_f, ω_f) to a real $x := ((a_n)_n, M)$ in I is defined to be

$$(h_f(a_n, n))_n$$

with modulus $\max(\alpha_f(k+2), M(\omega_f(k+1)-1))$. This is a modulus, for

$$\begin{aligned}&|h_f(a_m, m) - h_f(a_n, n)|\\ &\leqslant |h_f(a_m, m) - h_f(a_m, p)| + |h_f(a_m, p) - h_f(a_n, p)| + |h_f(a_n, p) - h_f(a_n, n)|\\ &\leqslant 2^{-k-2} + 2^{-k-1} + 2^{-k-2}\end{aligned}$$

if $m, n \geqslant M(\omega_f(k+1)-1)$ and $p \geqslant \alpha_f(k+1)$ (for the middle term), and moreover $m, n, p \geqslant \alpha_f(k+2)$ (for the first and last term). We denote this real by $f(x)$. The set of all such reals is called the *range* of f.

Two continuous functions $f\colon I \to \mathbb{R}$ and $g\colon J \to \mathbb{R}$ are *equal* if $I = J$ and $f(a) = g(a)$ for all rationals $a \in I$.

We show that indeed a continuous function f has ω_f as a modulus of uniform continuity.

Lemma (ContMod) *Let f be continuous and x, y be reals in its domain. Then*

$$|x-y| \leqslant 2^{-\omega_f(k)} \to |f(x) - f(y)| \leqslant 2^{-k}.$$

Proof Let $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$. Assume $|a_n - b_n| \leqslant 2^{-\omega_f(k)+1}$ for $n \geqslant p$. Then $|h_f(a_n, n) - h_f(b_n, n)| \leqslant 2^{-k}$ for $n \geqslant p, \alpha_f(k)$, that is $|f(x) - f(y)| \leqslant 2^{-k}$.

Essentially the same proof shows that application is compatible with equality.

Lemma (ContAppComp) *Let f be a continuous function and x, y be reals in its domain. Then $x = y$ implies $f(x) = f(y)$.*

Proof Let $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$, and assume $x = y$. Given k, pick p such that $|a_n - b_n| \leqslant 2^{-\omega_f(k)+1}$ for $n \geqslant p$. Then, as in the proof above, $|f(x) - f(y)| \leqslant 2^{-k}$. Hence $f(x) = f(y)$.

3.3 Continuous Functions and Limits

We show that continuous functions commute with limits.

Lemma (ContLim) *Let $(x_n)_n$ be a sequence of reals which converges to y. Assume $x_n, y \in I$ and let $f\colon I \to \mathbb{R}$ be continuous. Then $(f(x_n))_n$ converges to $f(y)$.*

Proof For a given k, pick p such that $p \leqslant n \to |x_n - y| \leqslant 2^{-\omega_f(k)}$ for all n. Then by the previous lemma $p \leqslant n \to |f(x_n) - f(y)| \leqslant 2^{-k}$. Hence $(f(x_n))_n$ converges to $f(y)$.

Lemma (ContRat) *Assume that $f, g\colon I \to \mathbb{R}$ are continuous and coincide on all rationals $a \in I$. Then $f = g$.*

Proof Let $x := ((a_n)_n, M)$. By ContLim, $(f(a_n))_n$ converges to $f(x)$ and $(g(a_n))_n$ to $g(x)$. Now $f(a_n) = g(a_n)$ implies $f(x) = g(x)$.

3.4 Intermediate Value Theorem

We next supply the standard constructive versions of the *intermediate value theorem.*

Theorem (Approximate Intermediate Value Theorem) *Let $a < b$ be rational numbers. For every continuous function $f\colon [a,b] \to \mathbb{R}$ with $f(a) \leqslant 0 \leqslant f(b)$, and every k, we can find $c \in [a,b]$ such that $|f(c)| \leqslant 2^{-k}$.*

Proof In the sequel we repeatedly invoke the Approximate Splitting Principle from Section 2.5. Given k, let $\varepsilon := 2^{-k}$. We 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 l so that, for all $x, y \in [a,b]$, if $|x - y| \leqslant 2^{-l}$, then $|f(x) - f(y)| \leqslant \varepsilon$, and divide $[a,b]$ into $a = a_0 < a_1 < \ldots < a_m = b$ such that $|a_{i-1} - a_i| \leqslant 2^{-l}$. Compare every $f(a_i)$ with $-\frac{\varepsilon}{2} < \frac{\varepsilon}{2}$. By assumption $f(a_0) < -\frac{\varepsilon}{2}$ and $\frac{\varepsilon}{2} < f(a_m)$; whence 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 $-\varepsilon < -\varepsilon/2$ and $f(a_{j+1})$ with $\varepsilon/2 < \varepsilon$. 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 $f(a_j) < -\varepsilon/2$ and $\varepsilon/2 < f(a_{j+1})$, then we would have $|f(a_{j+1}) - f(a_j)| > \varepsilon$, contradicting $|a_{j+1} - a_j| \leqslant 2^{-l}$.

Alternative Proof We give a different proof, which more directly makes use of the fact that our continuous functions come with witnessing data.

We may assume $f(a) < -2^{-k-1}$ and $2^{-k-1} < f(b)$ (see above). Divide $[a,b]$ into $a = a_0 < a_1 < \ldots < a_m = b$ such that $|a_{i-1} - a_i| \leqslant 2^{-\omega_f(k+1)}$. Consider all finitely many

$$h(a_i, n_0) \quad \text{for } i = 1, \rtimes ts, m,$$

with $n_0 := \alpha_f(k+1)$. Pick j such that $h(a_{j-1}, n_0) \leqslant 0 \leqslant h(a_j, n_0)$; this can be done because $f(a) < -2^{-k-1}$ and $2^{-k-1} < f(b)$. We show $|f(a_j)| \leqslant 2^{-k}$; for this it clearly suffices to show $|h(a_j, n)| \leqslant 2^{-k}$ for $n \geqslant n_0$. Now

$$|h(a_j, n)| \leqslant \left|h(a_j, n) - h(a_j, n_0)\right| + \left|h(a_j, n_0)\right| \leqslant 2^{-k-1} + 2^{-k-1},$$

where the first estimate holds by the choice of n_0, and the second one follows from the choice of a_j and $|h(a_{i-1}, n) - h(a_i, n)| \leqslant 2^{-k-1}$.

A problem with both of these proofs is that the algorithms they provide are rather bad: in each case one has to partition the interval into as many pieces as the modulus of the continuous function requires for the given error bound, and then for each of these (many) pieces perform certain operations. This problem seems to be unavoidable, since our continuous function may be rather flat. However, we can do somewhat better if we assume a uniform lower bound on the slope of f, that is, some $l \in \mathbb{N}$ such that for all $c, d \in \mathbb{Q}$ and all $m \in \mathbb{N}$

$$\frac{1}{2^m} \leqslant d - c \to f(c) <_{m+l} f(d).$$

Instead of the linear function $m \mapsto m + l$ one could use (as in [2]) an arbitrary "modulus of increase".

We begin with an auxiliary lemma for the Intermediate Value Theorem, which from a "correct" interval $c < d$ (that is, $f(c) \leqslant 0 \leqslant f(d)$ and $2^{-n} \leqslant d - c$) constructs a new one $c_1 < d_1$ with $d_1 - c_1 = \frac{2}{3}(d - c)$.

Lemma (IVTAux) *Let $f\colon [a, b] \to \mathbb{R}$ be continuous, and with a uniform lower bound l on its slope. Assume $a \leqslant c < d \leqslant b$, say $2^{-n} < d - c$, and $f(c) \leqslant 0 \leqslant f(d)$. Then we can construct c_1, d_1 with $d_1 - c_1 = 2/3(d - c)$, such that again $a \leqslant c \leqslant c_1 < d_1 \leqslant d \leqslant b$ and $f(c_1) \leqslant 0 \leqslant f(d_1)$.*

Proof Let $c_0 = 2c + d/3$ and $d_0 = c + 2d/3$. From $2^{-n} < d - c$ we obtain $2^{-n-2} \leqslant d_0 - c_0$, so $f(c_0) <_{n+2+l} f(d_0)$. Now compare 0 with this proper interval, using ApproxSplit. In the first case we have $0 \leqslant f(d_0)$; then let $c_1 = c$ and $d_1 = d_0$. In the second case we have $f(c_0) \leqslant 0$; then let $c_1 = c_0$ and $d_1 = d$.

Theorem (Intermediate Value Theorem) *If $f\colon [a, b] \to \mathbb{R}$ is continuous with $f(a) \leqslant 0 \leqslant f(b)$, and with a uniform lower bound on its slope, then we can find $x \in [a, b]$ such that $f(x) = 0$.*

Proof Iterating the construction in the auxiliary lemma IVTAux above, we construct two sequences $(c_n)_n$ and $(d_n)_n$ of rationals such that for all n

$$\begin{aligned}
&a = c_0 \leqslant c_1 \leqslant \ldots \leqslant c_n < d_n \leqslant \ldots \leqslant d_1 \leqslant d_0 = b,\\
&f(c_n) \leqslant 0 \leqslant f(d_n),\\
&d_n - c_n = \bigl(2/3\bigr)^n (b - a).
\end{aligned}$$

Let x, y be given by the Cauchy sequences $(c_n)_n$ and $(d_n)_n$ with the obvious modulus. As f is continuous, $f(x) = 0 = f(y)$ for the real number $x = y$.

Remark The proposition can also be proved for locally nonconstant functions. A function $f\colon [a, b] \to \mathbb{R}$ is *locally nonconstant* whenever if $a \leqslant a' < b' \leqslant b$ and c is an arbitrary real, then $f(x) \neq c$ for some real $x \in [a', b']$. Note that if f is continuous, then there also is a rational with that property. Strictly monotonic functions are clearly locally nonconstant, and so are nonconstant real polynomials.

3.5 Inverse Functions

We prove that every continuous function with a uniform lower bound on its slope has a continuous inverse. A constructive proof of this fact has been given by Mandelkern [8]. More recently, J. Berger [2] introduced a concept he called "exact representation of continuous functions", and based on this gave a construction converting an exact representation of an increasing function into an exact representation of its inverse. The proof below is based on our concept of a continuous function as a type-1 object.

Theorem *Let $f\colon [a,b] \to \mathbb{R}$ be continuous with a uniform lower bound on its slope. Let $f(a) \leqslant a' < b' \leqslant f(b)$. We can find a continuous $g\colon [a',b'] \to \mathbb{R}$ such that $f(g(y)) = y$ for every $y \in [a',b']$ and $g(f(x)) = x$ for every $x \in [a,b]$ such that $a' \leqslant f(x) \leqslant b'$.*

Proof Let $f\colon [a,b] \to \mathbb{R}$ be continuous with a uniform lower bound on its slope, that is, some $l \in \mathbb{N}$ such that for all $c, d \in [a,b]$ and all $m \in \mathbb{N}$,

$$2^{-m} \leqslant d - c \to f(c) <_{m+l} f(d).$$

Let $f(a) \leqslant a' < b' \leqslant f(b)$. We construct a continuous $g\colon [a',b'] \to \mathbb{R}$.

Let $u \in [a',b']$ be rational. Using $f(a) - u \leqslant a' - u \leqslant 0$ and $0 \leqslant b' - u \leqslant f(b) - u$, the Intermediate Value Theorem gives us an x such that $f(x) - u = 0$, as a Cauchy sequence (c_n). Let $h_g(u,n) := c_n$. Define the modulus α_g such that for $n \geqslant \alpha_g(k)$, $(2/3)^n(b-a) \leqslant 2^{-\omega_f(k+l+2)}$. For the uniform modulus ω_g of continuity assume $a' \leqslant u < v \leqslant b'$ and $k \in \mathbb{N}$. We claim that with $\omega_g(k) := k + l + 2$ (l from the hypothesis on the slope) we can prove the required property

$$|u - v| \leqslant 2^{-\omega_g(k)+1} \to |h_g(u,n) - h_g(v,n)| \leqslant 2^{-k} \quad (n \geqslant \alpha_g(k)).$$

Let $a' \leqslant u < v \leqslant b'$ and $n \geqslant \alpha_g(k)$. For $c_n^{(u)} := h_g(u,n)$ and $c_n^{(v)} := h_g(v,n)$ assume that $|c_n^{(u)} - c_n^{(v)}| > 2^{-k}$; we must show $|u - v| > 2^{-\omega_g(k)+1}$.

By the proof of the Intermediate Value Theorem we have

$$d_n^{(u)} - c_n^{(u)} \leqslant (2/3)^n(b-a) \leqslant 2^{-\omega_f(k+l+2)} \quad \text{for } n \geqslant \alpha_g(k).$$

Using $f(c_n^{(u)}) - u \leqslant 0 \leqslant f(d_n^{(u)}) - u$, ContMod in Section 3.2 gives us

$$|f(c_n^{(u)}) - u| \leqslant |(f(d_n^{(u)}) - u) - (f(c_n^{(u)}) - u)| = |f(d_n^{(u)}) - f(c_n^{(u)})| \leqslant 2^{-k-l-2}$$

and similarly $|f(c_n^{(v)}) - v| \leqslant 2^{-k-l-2}$. Hence, using $|f(c_n^{(u)}) - f(c_n^{(v)})| \geqslant 2^{-k-l}$ (which follows from $|c_n^{(u)} - c_n^{(v)}| > 2^{-k}$ by the hypothesis on the slope),

$$|u - v| \geqslant |f(c_n^{(u)}) - f(c_n^{(v)})| - |f(c_n^{(u)}) - u| - |f(c_n^{(v)}) - v| \geqslant 2^{-k-l-1}.$$

Now $f(g(u)) = u$ follows from

$$|f(g(u)) - u| = |h_f(c_n, n) - u| \leqslant |h_f(c_n, n) - h_f(c_n, m)| + |h_f(c_n, m) - u|,$$

which is $\leqslant 2^{-k}$ for $n, m \geqslant \alpha_f(k+1)$. By ContRat, $f(g(y)) = y$ for $y \in [a', b']$.

For every $x \in [a, b]$ with $a' \leqslant f(x) \leqslant b'$, from $g(f(x)) < x$ we obtain the contradiction $f(x) = f(g(f(x))) < f(x)$ by the hypothesis on the slope, and similarly for $>$. So we must have $g(f(x)) = x$.

As an example, consider the squaring function $f\colon [1, 2] \to [1, 4]$, given by the approximating map $h_f(a, n) := a^2$, constant Cauchy modulus $\alpha_f(k) := 1$, modulus $\omega_f(k) := k + 1$ uniform continuity and lower and upper bounds $N_f = 1$ and $M_f = 4$. The lower bound on its slope is $l := 0$, because for all $c, d \in [1, 2]$

$$2^{-m} \leqslant d - c \to c^2 <_m d^2.$$

Then $h_g(u, n) := c_n^{(u)}$, as constructed in the IVT for $x^2 - u$, iterating IVTAux. The Cauchy modulus α_g is such that $(2/3)^n \leqslant 2^{-k+3}$ for $n \geqslant \alpha_g(k)$, and the modulus of uniform continuity is $\omega_f(k) := k + 2$.

4 Program Extraction

We now address the issue of extracting computational content of some of these proofs, with machine help. Formalization clearly involves a lot of details, beginning with an appropriate representation of the data. Here we take a direct approach, by explicitly building the required number systems (natural numbers in binary, rationals, reals as Cauchy sequences of rationals with a modulus, continuous functions in the sense of the type-1 representation described above, etc.) It turns out that in this setup extracted programs perform reaonably well.

4.1 Some Background on Program Extraction

Before describing details of the extracted programs we provide some proof-theoretic background on program extraction and its optimization in general and on problems arising in this case study in particular.

The method of program extraction used in this paper is based on *modified realizability* as introduced by Kreisel [7]. In short, from every constructive proof M of a non-Harrop formula A (in natural deduction or a similar proof calculus) one extracts a program $[\![M]\!]$ "realizing" A, essentially, by removing computationally irrelevant parts from the proof (proofs of Harrop formulas have no computational content). The extracted program has some simple type $\tau(A)$ which depends on the logical shape of the proven formula A only. In its original form the extraction process is fairly straightforward, but usually leads to unnecessarily complex programs. In order to obtain better programs, proof assistants (for instance Coq, Isabelle/HOL, Agda, Nuprl, Minlog) offer various optimizations of program extraction. Below we describe some optimizations implemented in Minlog [10], which are relevant for our present case study.

Animation Suppose a proof of a theorem uses a lemma. Then the proof term contains just the name of the lemma, say `L`. In the term extracted from this proof we want to preserve the structure of the original proof as much as possible, and hence we use a new constant `cL` at those places where the computational content of the lemma is needed. When we want to execute the program, we have to replace the constant `cL` corresponding to a lemma `L` by the extracted program of its proof. This can be achieved by adding computation rules for `cL` and `cGA`. We can be rather flexible here and enable/block rewriting by using animate/deanimate as desired.

Let It often happens that a subterm has many occurrences in a term, which leads to unwanted recomputations when evaluating it. A possible cure is to "optimize" the term after extraction, and replace for instance $M[x := N]$ with many occurrences of x in M by $(\lambda x M)N$ (or a corresponding "let"-expression). However, this can already be done at the proof level: When an object (value of a variable or realizer of a premise) might be used more than once, make sure (if necessary by a cut) that the goal has the form $A \to B$ or $\forall_x A$. Now use the "identity lemma" `Id`: $\hat{P} \to \hat{P}$, with a predicate variable $\hat{P}$. Its realizer then has the form $\lambda f, x. fx$. If `cId` is not animated, the extracted term has the form `cId`$(\lambda x M)N$, which is printed as [`let` x N M].

Quantifiers without computational content Besides the usual quantifiers, $\forall$ and $\exists$, Minlog has so-called *non-computational quantifiers*, $\forall^{\mathsf{nc}}$ and $\exists^{\mathsf{nc}}$, which allow for the extraction of simpler programs. The nc-quantifiers, which were first introduced in [3], can be viewed as a refinement of the Set/Prop distinction in constructive type systems like Coq or Agda. Intuitively, a proof of $\forall_x^{\mathsf{nc}} A(x)$ ($A(x)$ non-Harrop) represents a procedure that assigns to every x a proof $M(x)$ of $A(x)$ where $M(x)$ does not make "computational use" of x, i.e., the extracted program $[\![M(x)]\!]$ does not depend on x. Dually, a proof of $\exists_x^{\mathsf{nc}} A(x)$ is proof of $M(x)$ for some x where the witness x is "hidden", that is, not available for computational use. Consequently, the types of extracted programs for nc-quantifiers are $\tau(\forall_{x^\rho}^{\mathsf{nc}} A) = \tau(\exists_{x^\rho}^{\mathsf{nc}} A) = \tau(A)$ as opposed to $\tau(\forall_{x^\rho} A) = \rho \Rightarrow \tau(A)$ and $\tau(\exists_{x^\rho} A) = \rho \times \tau(A)$. The extraction rules are, for example in the case of $\forall^{\mathsf{nc}}$-introduction and -elimination, $[\![(\lambda x. M^{A(x)})^{\forall_x^{\mathsf{nc}} A(x)}]\!] = [\![M]\!]$ and $[\![(M^{\forall_x^{\mathsf{nc}} A(x)} t)^{A(t)}]\!] = [\![M]\!]$ as opposed to $[\![(\lambda x. M^{A(x)})^{\forall_x A(x)}]\!] = [\![\lambda x M]\!]$ and $[\![(M^{\forall_x A(x)} t)^{A(t)}]\!] = [\![Mt]\!]$. In order for the extracted programs to be correct the variable condition for $\forall^{\mathsf{nc}}$-introduction needs to be strengthened by requiring in addition the abstracted variable x not to occur in the extracted program $[\![M]\!]$. Note that for a Harrop formula A the formulas $\forall_x^{\mathsf{nc}} A$ and $\forall_x A$ are equivalent. Similarly, $\exists_x^{\mathsf{nc}} A$ and $\exists_x A$ are equivalent.

Rules Constants – denoting partial computable functionals – are defined via (computation and rewrite) rules (cf. [11] for details). To simplify equational reasoning, the system identifies terms with the same "normal form", and we should be able to add rewrite rules used to generate normal forms. Decidable predicates are viewed as boolean valued functions, so that the rewrite mechanism applies to them as well.

4.2 Program Extraction for the Intermediate Value Theorem

Since the formalized proofs follow rather closely the informal development above, we do not comment on how these proofs are built, but only on the extracted terms.

Here is the extracted term for `ApproxSplit`.

```
(Rec real=>real=>real=>pos=>boole) ([as4,M5]
 (Rec real=>real=>pos=>boole)
 ([as9,M10]
   (Rec real=>pos=>boole)
   ([as13,M14,n15]
     as13(M5(S(S n15))max M10(S(S n15))max M14(S(S n15)))<=
     (as4(M5(S(S n15))max M10(S(S n15)))+
      as9(M5(S(S n15))max M10(S(S n15))))/2)))
```

of type `real=>real=>real=>pos=>boole`. It takes three reals given x, y, z with moduli M, N, K (here given by their Cauchy sequences `as4`, `as9`, `as13` and moduli `M5`, `M10`, `M14`) and a positive number k (here `n15`), and computes $p := \max(M(k+2), N(k+2))$ and $q := \max(p, L(k+2))$. Then the choice whether to go right or left is by computing the boolean value $c_q \leqslant a_p + b_p/2$.

For the auxiliary lemma `IVTAux` we obtain the extracted term

```
[f0,n1,n2]
 (cId rat@@rat=>rat@@rat)
 ([cd4]
   [let cd5
     ((2#3)*left cd4+(1#3)*right cd4@
      (1#3)*left cd4+(2#3)*right cd4)
     [if (cApproxSplit(RealConstr(f0 approx left cd5)
                                 ([n6]f0 uMod(S(S n6))))
                      (RealConstr(f0 approx right cd5)
                                 ([n6]f0 uMod(S(S n6))))
          0
          (S(S(n2+n1))))
      (left cd4@right cd5)
      (left cd5@right cd4)]])
```

of type `cont=>pos=>pos=>rat@@rat=>rat@@rat`. As in the informal proof, it takes a continuous f (here `f0`), a uniform lower bound l on its slope (here `n1`), a positive number n (here `n2`) and two rationals c, d (here the pair `cd4`) such that $2^{-n} < d - c$. Let $c_0 := 2c + d/3$ and $d_0 := c + 2d/3$ (here the pair `cd5`, introduced via `let` because it is used four times). Then `ApproxSplit` is applied to $f(c_0)$, $f(d_0)$, 0 and the witness $n + 2 + l$ (here `S(S(n2+n1))`) for $f(c_0) < f(d_0)$. In the first case we go left, that is $c_1 := c$ and $d_1 := d_0$, and in the second case we go right, that is $c_1 := c_0$ and $d_1 := d$.

In the proof of the Intermediate Value Theorem, the construction step in IVTAux (from a pair c, d to the "better" pair c_0, d_0) had to be iterated, to produce two sequences $(c_n)_n$ and $(d_n)_n$ of rationals. This is the content of a separate lemma IVTcds, whose extracted term is

```
[f0,n1,n2](cDC rat@@rat)(f0 doml@f0 domr)
                        ([n4]cIVTAux f0 n1(n2+n4))
```

of type cont=>pos=>pos=>pos=>rat@@rat. It takes a continuous $f\colon [a, b] \to \mathbb{R}$ (here f0), a uniform lower bound l on its slope (here n1), and a positive number k_0 (here n2) such that $2^{-k_0} < b - a$. Then the axiom of dependent choice DC is used, to construct from an initial pair $(c_0, d_0) = (a, b)$ of rationals (here f0 doml@f0 domr) a sequence of pairs of rationals, by iterating the computational content cIVTAux of the lemma IVTAux.

Now we are ready to formalize the proof of the Intermediate Value Theorem IVTFinal. Its extracted term is

```
[f0,n1,n2,n3]RealConstr([n4]left(cIVTcds f0 n1 n2 n4))
                       ([n4]SZero(S(n4+n3)))
```

of type cont=>pos=>pos=>pos=>real. It takes a continuous f (here f0), a uniform lower bound l on its slope (here n1), a positive number k_0 (here n2) such that $2^{-k_0} < b - a$, and a positive number k_1 (here n3) such that $b - a < 2^{k_1}$. Then the desired zero x of f is constructed as the Cauchy sequence $(c_n)_n$, with a modulus depending on k_1.

To compute approximations of $\sqrt{2}$ we need RealApprox, stating that every real can be approximated by a rational. Its extracted term is

```
(Rec real=>pos=>rat)([as2,M3,n4]as2(M3 n4))
```

of type real=>pos=>rat. It takes a real x (here given by the Cauchy sequence as2 and modulus M3) and a positive number k (here n4), and computes a rational a such that $|x - a| \leqslant 2^{-k}$.

In order to compose IVTFinal with RealApprox, we prove a proposition IVTApprox stating that given an error bound, we can find a rational approximating $\sqrt{2}$ up to this bound. Clearly we need to refer to $\sqrt{2}$ in the statement of the theorem, but on the other hand we do not want to see a representation of $\sqrt{2}$ in the extracted term, but only the construction of the rational approximation from the error bound. Therefore in the statement of IVTApprox we use the non-computational quantifier $\exists^{\mathrm{nc}}$, for the zero z of the given continuous f. The extracted term of IVTApprox then simply is

```
[f0,n1,n2,n3]cRealApprox(cIVTFinal f0 n1 n2 n3)
```

of type cont=>pos=>pos=>pos=>pos=>rat.

Now we "animate" the auxiliary lemmas, that is, unfold all constants with "`c`" in front of name of the lemma. For `IVTApprox` this gives

```
[f0,n1,n2,n3,n4]
 left
 [let cd5
   ((cDC rat@@rat)(f0 doml@f0 domr)
   ([n5]
     (cId rat@@rat=>rat@@rat)
     ([cd7]
       [let cd8
         ((2#3)*left cd7+(1#3)*right cd7@
          (1#3)*left cd7+(2#3)*right cd7)
         [if (0<=(f0 approx left cd8
                  (f0 uMod(S(S(S(S(S(S(n2+n5+n1)))))))+
                 f0 approx right cd8
                  (f0 uMod(S(S(S(S(S(S(n2+n5+n1))))))))/2)
          (left cd7@right cd8)
          (left cd8@right cd7)]]))
   (PosPred(SZero(S(n4+n3)))))
   [let cd6
    ((2#3)*left cd5+(1#3)*right cd5@
     (1#3)*left cd5+(2#3)*right cd5)
    [if
     (0<=
      (f0 approx left cd6
       (f0 uMod(S(S(S(S(S(S(S(S(n2+n4+n3+n4+n3+n1)))))))))+
       f0 approx right cd6
       (f0 uMod(S(S(S(S(S(S(S(S(n2+n4+n3+n4+n3+n1))))))))))/2)
     (left cd5@right cd6)
     (left cd6@right cd5)]]]
```

Let us now use this term to compute approximations of $\sqrt{2}$. First we construct the continuous function $x \mapsto x^2-2$ on $[1, 2]$, with its (trivial) uniform Cauchy modulus and modulus of uniform continuity, and give it the name `a-sq-minus-two`:

```
(define a-sq-minus-two
  (pt "contConstr 1 2([a0,n1]a0*a0-2)([n0]1)S"))
```

We now apply the extracted term of theorem `IVTApprox` to `a-sq-minus-two` and in addition a uniform lower bound l on its slope, a positive number k_0 such that $2^{-k_0} < b - a$, and a positive number k_1 such that $b - a < 2^{k_1}$ (which all happen to be 1 in this case), and normalize the result:

```
(define sqrt-two-approx
  (normalize-term
    (apply mk-term-in-app-form
      (list (proof-to-extracted-term
```

```
                (theorem-name-to-proof "IVTApprox"))
             a-sq-minus-two  (pt "1") (pt "1") (pt "1")))))
```

which prints as

```
[n0]
 left[let cd1
       ((cDC rat@@rat)(1@2)
       ([n1]
         (cId rat@@rat=>rat@@rat)
         ([cd3]
           [let cd4
             ((2#3)*left cd3+(1#3)*right cd3@
              (1#3)*left cd3+(2#3)*right cd3)
             [if (0<=(left cd4*left cd4-2+
                     (right cd4*right cd4-2))/2)
              (left cd3@right cd4)
              (left cd4@right cd3)]]))
       (PosPred(SZero(S(S n0)))))
       [let cd2
        ((2#3)*left cd1+(1#3)*right cd1@
         (1#3)*left cd1+(2#3)*right cd1)
        [if (0<=(left cd2*left cd2-2+
                (right cd2*right cd2-2))/2)
         (left cd1@right cd2)
         (left cd2@right cd1)]]]
```

The term `sqrt-two-approx` has type `pos=>rat`, where the argument k is for the error bound 2^{-k}. We can now directly (that is, without first translating it into a programming language) use it to compute an approximation of $\sqrt{2}$ to say 20 binary digits. To do this, we need to "animate" `Id` and then normalize the result of applying `sqrt-two-approx` to 20 (we use normalization by evaluation here, for efficiency reasons):

```
(animate "Id")
(pp (nbe-normalize-term-without-eta
      (make-term-in-app-form sqrt-two-approx (pt "20"))))
```

The result (returned in 1.5 seconds) is the rational

```
464225451859201404985#328256967394537077627
```

or 1.4142135520957329, which differs from $\sqrt{2} = 1.4142135623730951$ at the seventh (decimal) digit.

It should be noted that for efficiency reasons we have equipped our constants for arithmetical operations $+, *, -, /, <$ on rationals with "external code". This means that when the arguments are numerals already, then the result is computed with the (much faster) arithmetic function of the programming language, not by the rules defining the internal constants.

For a further speed-up, we can also translate this internal term (where "internal" means "in our underlying logical language", hence usable in formal proofs) into an expression of a programming language (Scheme in our case). This done by evaluating (term-to-expr sqrt-two-approx):

```
(lambda (n0)
  (car
   (let ([cd1
          (((cdc (cons 1 2))
            (lambda (n1)
              (lambda (cd3)
                (let ([cd4
                       (cons (+ (* 2/3 (car cd3))
                                (* 1/3 (cdr cd3)))
                             (+ (* 1/3 (car cd3))
                                (* 2/3 (cdr cd3))))])
                  (if (<= 0
                          (/ (+ (- (* (car cd4) (car cd4)) 2)
                                (- (* (cdr cd4) (cdr cd4)) 2))
                             2))
                      (cons (car cd3) (cdr cd4))
                      (cons (car cd4) (cdr cd3)))))))
           (pospred (* (+ (+ n0 1) 1) 2)))])
     (let ([cd2
            (cons (+ (* 2/3 (car cd1)) (* 1/3 (cdr cd1)))
                  (+ (* 1/3 (car cd1)) (* 2/3 (cdr cd1))))])
       (if (<= 0
               (/ (+ (- (* (car cd2) (car cd2)) 2)
                     (- (* (cdr cd2) (cdr cd2)) 2))
                  2))
           (cons (car cd1) (cdr cd2))
           (cons (car cd2) (cdr cd1)))))))
```

The result is very close to the internal term displayed above; we have replaced the internal constant cDC (computational content of the axiom of dependent choice) by the corresponding Scheme function (a curried form of iteration):

```
(define cdc
  (lambda (init)
    (lambda (step)
      (lambda (n)
        (if (= 1 n)
            init
            ((step n) (((cdc init) step) (- n 1)))))))),
```

the internal arithmetical functions `+`, `*`, `/` by the ones from the programming language and the internal pairing and unpairing functions by `cons`, `car` and `cdr`. – It turns out that this code is reasonably fast: evaluating

```
((ev (term-to-expr sqrt-two-approx)) 20))
```

gives the result in 10 ms.

References

1. Patrik Andersson. Exact real arithmetic with automatic error estimates in a computer algebra system. Master's thesis, Mathematics department, Uppsala University, 2001.
2. Josef Berger. Exact calculation of inverse functions. *Math. Log. Quart.*, 51(2):201–205, 2005.
3. Ulrich Berger. Program extraction from normalization proofs. In M. Bezem and J.F. Groote, editors, *Typed Lambda Calculi and Applications*, vol. 664 of *LNCS*, pp. 91–106. Springer Verlag, Berlin, Heidelberg, New York, 1993.
4. Errett Bishop. *Foundations of Constructive Analysis*. McGraw-Hill, New York, 1967.
5. Errett Bishop. Mathematics as a numerical language. In J. Myhill A. Kino and R.E. Vesley, editors, *Intuitionism and Proof Theory, Proceedings of the summer conference at Buffalo N.Y. 1968*, Studies in logic and the foundations of mathematics, pp. 53–71. North–Holland, Amsterdam, 1970.
6. Luis Cruz-Filipe. *Constructive Real Analysis: A Type-Theoretical Formalization and Applications*. PhD thesis, Nijmegen University, 2004.
7. Georg Kreisel. Interpretation of analysis by means of constructive functionals of finite types. In A. Heyting, editor, *Constructivity in Mathematics*, pp. 101–128. North–Holland, Amsterdam, 1959.
8. Mark Mandelkern. Continuity of monotone functions. *Pacific J. Math.*, 99(2):413–418, 1982.
9. Erik Palmgren. Constructive nonstandard analysis. In A. Petry, editor, *Méthods et analyse non standard*, vol. 9, pp. 69–97. Cahiers du Centre de Logique, 1996.
10. Helmut Schwichtenberg. Minlog. In F. Wiedijk, editor, *The Seventeen Provers of the World*, vol. 3600 of *LNAI*, pp. 151–157. Springer Verlag, Berlin, 2006.
11. Helmut Schwichtenberg. Recursion on the partial continuous functionals. In C. Dimitracopoulos, L. Newelski, D. Normann and J. Steel, editors, Logic Colloquium 2005, vol. 26 of Lecture Notes in Logic, pp. 173–201. Amer. Math. Soc., 2006.

Brouwer's Approximate Fixed-Point Theorem is Equivalent to Brouwer's Fan Theorem

Wim Veldman

Μὴ γνώτω ἡ ἀριστερά σου τί ποιεῖ ἡ δεξιά σου
Do not let your left hand know what your right is doing
Matthew 6,3

Abstract In a weak system for intuitionistic analysis, one may prove, using the Fan Theorem as an additional axiom, that, for every continuous function ϕ from the unit square U to itself, for every positive rational e, there exists x in U such that $|\phi(x) - x| < e$. Conversely, if this statement is taken as an additional axiom, the Fan Theorem follows.

1 Introduction

L.E.J. Brouwer's fame as a mathematician rests on two memorable feats.

First, he discovered, in 1911, see [2], the Dimension Theorem, saying that, for all m, n, if $[0, 1]^m$ is homeomorphic to $[0, 1]^n$, then $m = n$, and his closely related Fixed-Point Theorem saying that, for each n, every continuous function from $[0, 1]^n$ to $[0, 1]^n$ has a fixed point. By proving these results he founded the subject of algebraic topology.

Secondly, reflecting on the meaning of mathematical statements, he saw the need for a reform of the existing mathematical practice and started *intuitionistic mathematics*.

He observed that his Fixed-Point Theorem is not constructively valid and published, in 1952, see [4] and [5], an intuitionistically correct approximate version. Long before, in 1924, see [3], when trying to prove that every real function from $[0, 1]$ to $\mathbb{R}$ is uniformly continuous, he had formulated the Fan Theorem. He "*proved*" the Fan Theorem by a much-debated philosophical argument that, in fact, has some consequences much stronger than the Fan Theorem, see [14] and [15].

We want to establish the noteworthy fact that the Approximate Fixed-Point Theorem and the Fan Theorem, icons for the two pillars of Brouwer's fame, when we consider them in the spirit of the Reverse Mathematics Program initiated by H. Friedman and S.G. Simpson, turn out to be equivalent.

W. Veldman (✉)
Institute for Mathematics, Astrophysics and Particle Physics, Faculty of Science, Radboud University Nijmegen, Postbus 9044, 6500 KD Nijmegen, the Netherlands
e-mail: W.Veldman@science.ru.nl

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341, DOI 10.1007/978-1-4020-8926-8_14, © Springer Science+Business Media B.V. 2009

This fact became clear after some reflection on the recursive counterexample to Brouwer's Fixed-Point Theorem due to V.P. Orevkov, see [10] and also [11] and [12].

Some of the results of this paper also occur in [13], but the treatment here is slightly different.

The one-dimensional case of Brouwer's Fixed-Point Theorem is closely related to the Intermediate Value Theorem. In Section 4, we show that the exact version of the one-dimensional fixed-point theorem fails constructively, and that the approximate version is true.

In the last theorem of the paper, we prove that the Fan Theorem is also equivalent to the statement that every continuous function from $[0, 1]^2$ to $[0, 1]^2$ with at most one fixed point has a fixed point. It follows that the latter statement is intuitionistically true.

I want to express my thanks to Ruben van den Brink for spotting several inaccuracies in an earlier version of this paper.

2 The Formal Context

We introduce a formal system, BIM, for Basic Intuitionistic Mathematics.

There are two kinds of variables, *numerical* variables $m, n, p, \ldots$, whose intended range is the set $\mathbb{N}$ of the natural numbers, and *function* variables $\alpha, \beta, \gamma, \ldots$, whose intended range is the set $\mathcal{N}$ of all infinite sequences of natural numbers, that is, functions from $\mathbb{N}$ to $\mathbb{N}$. For every α, for every n, $\alpha(n)$ is the value that α assumes at n. There is a numerical constant 0. There are are unary function constants $\underline{0}$, a name for the function with the constant value 0, and S, a name for the successor function, and K, L, names for the projection functions. There is one binary function symbol J, a name for the pairing function. From these symbols *numerical terms* are formed in the usual way. The basic terms are the numerical variables and the numerical constant and more generally, a term is obtained from earlier constructed terms by the use of a function symbol. Function variables are at this stage the only *function terms*. As the theory develops, names for operations on infinite sequences will be introduced and more complicated function terms will appear.

There are two equality symbols, $=_0$ and $=_1$. The first symbol may be placed between numerical terms only and the second one between function terms only. When confusion seems improbable we simply write $=$ and not $=_0$ or $=_1$. A *basic formula* is an equality between numerical terms or an equality between function terms. A *basic formula in the strict sense* is an equality between numerical terms. We obtain the formulas of the theory from the basic formulas by using the connectives, the numerical quantifiers and the function quantifiers. The logic of the theory is of course intuitionistic logic.

We adopt the following **Axiom of Extensionality**:

$$\forall \alpha \beta [\alpha =_1 \beta \leftrightarrow \forall n [\alpha(n) =_0 \beta(n)]]$$

The Axiom of Extensionality guarantees that every formula will be provably equivalent to a formula built up by means of connectives and quantifiers from basic formulas in the strict sense.

We also need the following **Axioms on the function constants**:

$$\forall n[\neg(S(n) = 0)],\ \forall m\forall n[S(m) = S(n) \to m = n],\ \forall n[\underline{0}(n) = 0],$$
$$\forall m\forall n[K(J(m,n)) = m \wedge L(J(m,n)) = n]$$

Thanks to the presence of the pairing function we may treat binary, ternary and other non-unary operations on $\mathbb{N}$ as unary functions. "$\alpha(m, n, p)$" for instance will be an abbreviation of "$\alpha(J(J(m,n), p))$".

Next we introduce **Axioms on the closure of the universe of functions under the recursive operations**:

Composition:
$\forall\alpha\forall\beta\exists\gamma\forall n[\gamma(n) = \alpha(\beta(n))]$
Primitive Recursion:
$\forall\alpha\forall\beta\exists\gamma\forall m\forall n[\gamma(m,0) = \alpha(m) \wedge \gamma(m, S(n)) = \beta(m, n, \gamma(m,n))]$
Unbounded Search:
$\forall\alpha[\forall m\exists n[\alpha(m,n) = 0] \to \exists\gamma\forall m[\alpha(m, \gamma(m)) = 0]]$.

We sometimes call the Axiom of Unbounded Search the *Minimal Axiom of Countable Choice*.

We also add the **Axiom Scheme of Induction**:

For every formula $\phi = \phi(n)$ the universal closure of the following formula is an axiom:

$$(\phi(0) \wedge \forall n[\phi(n) \to \phi(S(n))]) \to \forall n[\phi(n)]$$

The system consisting of the axioms mentioned up to now will be called BIM.

We assume that constants for the primitive recursive functions and relations with their defining equations have been added to BIM. In particular, we assume that some primitive recursive coding of finite sequences of natural numbers by natural numbers is given:

$$(m_0, \ldots, m_{k-1}) \mapsto \langle m_0, \ldots, m_{k-1}\rangle$$

Every natural number is supposed to code exactly one finite sequence of natural numbers. For each a we let *length(a)* be the length of the finite sequence coded by a. For each a, for each $i < length(a)$, we let $a(i)$ be the value of the finite sequence coded by a at i. We let $*$ denote the binary function corresponding to concatenation of finite sequences, so, for each a, b, $a * b$ is the number coding the finite sequence that we obtain by putting the finite sequence coded by b behind the finite sequence coded by a.

For each a, for each $n \leqslant length(a)$, we define: $\overline{a}(n) = \langle a(0), \ldots, a(n-1)\rangle$. If confusion seems unlikely, we sometimes write: "$\overline{a}n$" and not: "$\overline{a}(n)$".

For all a, b we define: *a is an initial segment of* b, notation: $a \sqsubseteq b$, if and only if there exists $n \leqslant length(b)$ such that $a = \overline{b}n$, and: *a is a proper initial part of b* if and only if both $a \sqsubseteq b$ and $a \neq b$.

For each α, for each n, we define $\overline{\alpha}(n) = \langle \alpha(0), \ldots \alpha(n-1)\rangle$. If confusion seems unlikely, we sometimes write: "$\overline{\alpha}n$" and not: "$\overline{\alpha}(n)$".

We use the letter $\mathcal{C}$ in order to denote *Cantor space*, so "$\alpha \in \mathcal{C}$" is an abbreviation for "$\forall n[\alpha(n) = 0 \vee \alpha(n) = 1]$".

We now are able to formulate the **Weak Fan Theorem**:

$$\forall\beta[\forall\alpha \in \mathcal{C}\exists n[\beta(\overline{\alpha}n) = 1] \rightarrow \exists m\forall\alpha \in \mathcal{C}\exists n[\beta(\overline{\alpha}n) = 1 \wedge \overline{\alpha}n \leqslant m]]$$

Let β belong to $\mathcal{C}$. We call the set of all n such that $\beta(n) = 1$ the subset of $\mathbb{N}$ *decided by* β, notation: D_β. A subset X of $\mathbb{N}$ is called *decidable* if and only if there exists β in $\mathcal{C}$ such that $X = D_\beta$.

Let s belong to $\mathbb{N}$. We call the set of all n such that $n < length(s)$ and $s(n) \neq 0$ the subset of $\mathbb{N}$ *decided by* s, notation: D_s. Note that, for every s, D_s is a finite subset of $\mathbb{N}$. Also note that, for every β in $\mathcal{C}$, $D_\beta = \bigcup_{n \in \mathbb{N}} D_{\overline{\beta}n}$.

Let Y be a subset of $\mathcal{N}$ and let X be a subset of $\mathbb{N}$. X is called a *bar in* Y if and only if, for each α in Y, there exists n such that $\overline{\alpha}n$ belongs to X.

The Weak Fan Theorem thus says:

Every decidable subset of $\mathbb{N}$ *that is a bar in Cantor space has a finite subset that is a bar in Cantor space*

Brouwer also believed the Weak Fan Theorem to be true in case we have a subset of $\mathbb{N}$ that is a bar in Cantor space but that we do not know to be decidable, see [13] and [14].

Moreover, Brouwer did not restrict himself to Cantor space. Cantor space is just one example of a *fan*. A subset $\mathcal{F}$ of $\mathcal{N}$ is called a *fan* if and only if it is *sequentially closed*, (that is, for each α, if, for each n, $\overline{\alpha}n$ contains an element of $\mathcal{F}$, then α belongs to $\mathcal{F}$), and the set of all s such that s contains an element of $\mathcal{F}$ is a decidable subset of $\mathbb{N}$ and, for each s, there exists m such that for all $i > m$, $s * \langle i\rangle$ does not contain an element of $\mathcal{F}$. The more general Fan Theorem, claiming that, for every fan $\mathcal{F}$, every decidable subset of $\mathbb{N}$ that is a bar in $\mathcal{F}$ has a finite subset that is a bar in $\mathcal{F}$, may be derived from the Weak Fan Theorem if we assume a weak axiom of Countable Choice, see [13].

We also want to study **Kleene's Alternative (to the Weak Fan Theorem)**:

$$\exists\beta[\forall\alpha \in \mathcal{C}\exists n[\beta(\overline{\alpha}n) = 1] \wedge \forall m\exists a[length(a) = m \wedge \forall i \leqslant m[\beta(\overline{a}i) \neq 1]]]$$

that is:

There exists a decidable subset of $\mathbb{N}$ *that is a bar in Cantor space, while every one of its finite subsets positively fails to be a bar in Cantor space*

The formal systems that we obtain from BIM, by adding the Weak Fan Theorem, or Kleene's Alternative, respectively, as an axiom, will be called WFT, KA, respectively.

The intuitionistic mathematician has a natural model both for the formal system WFT, and for the formal system KA. His own intuitionistic continuum is a model for WFT and the collection of the computable functions is one for KA, as has been shown by S.C. Kleene, see [9].

3 Uniform Continuity

We assume that a primitive recursive subset $\mathbb{Q}$ of $\mathbb{N}$ has been indicated and that a primitive recursive relation $<_{\mathbb{Q}}$ has been defined on $\mathbb{N}$, and that primitive recursive binary operations $+_{\mathbb{Q}}$, $-_{\mathbb{Q}}$ and $\cdot_{\mathbb{Q}}$ have been defined on $\mathbb{N}$ and special elements $0_{\mathbb{Q}}$ and $1_{\mathbb{Q}}$ of $\mathbb{Q}$ have been designated in such a way that $(\mathbb{Q}, <_{\mathbb{Q}}, +_{\mathbb{Q}}, -_{\mathbb{Q}}, \cdot_{\mathbb{Q}}, 0_{\mathbb{Q}}, 1_{\mathbb{Q}})$ is isomorphic to the ordered field of the rationals. We will call the elements of $\mathbb{Q}$, somewhat informally, by their usual names, like $\frac{2}{3}$, $-\frac{1}{2}$, and so on.

We let the set $\mathbb{S}$ of *non-degenerate rational segments* be the set of all pairs $\langle p, q\rangle$ such that p, q belong to $\mathbb{Q}$ and $p <_{\mathbb{Q}} q$.

We define binary relations $\#_{\mathbb{S}}$, $\approx_{\mathbb{S}}$, $\sqsubset_{\mathbb{S}}$ and $\sqsubseteq_{\mathbb{S}}$ on $\mathbb{N}$ such that for all p, q, r, s in $\mathbb{Q}$ with the property $p <_{\mathbb{Q}} q$ and $r <_{\mathbb{Q}} s$,

$\langle p, q\rangle \#_{\mathbb{S}} \langle r, s\rangle$ ($\langle p, q\rangle$ *lies apart from* $\langle r, s\rangle$) if and only if either $q <_{\mathbb{Q}} r$ or $s <_{\mathbb{Q}} p$,
$\langle p, q\rangle \approx_{\mathbb{S}} \langle r, s\rangle$ ($\langle p, q\rangle$ *does not lie apart from, or: touches, or: partially covers* $\langle r, s\rangle$) if and only if both $p \leqslant_{\mathbb{Q}} s$ and $r \leqslant_{\mathbb{Q}} q$,
$\langle p, q\rangle \sqsubset_{\mathbb{S}} \langle r, s\rangle$ ($\langle p, q\rangle$ *is strictly included in* $\langle r, s\rangle$) if and only if $r <_{\mathbb{Q}} p$ and $q <_{\mathbb{Q}} s$,
and $\langle p, q\rangle \sqsubseteq_{\mathbb{S}} \langle r, s\rangle$ ($\langle p, q\rangle$ *is included in* $\langle r, s\rangle$) if and only if $r \leqslant_{\mathbb{Q}} p$ and $q \leqslant_{\mathbb{Q}} s$.

For each p, q in $\mathbb{Q}$ such that $p \leqslant_{\mathbb{Q}} q$ we denote the number $q -_{\mathbb{Q}} p$ by $length_{\mathbb{S}}(\langle p, q\rangle)$.

Note that $\mathbb{Q}$ and $\mathbb{S}$ are subsets of $\mathbb{N}$.

Let α belong to $\mathcal{N}$. α is *a real number* if and only if (i) for each n, $\alpha(n)$ belongs to $\mathbb{S}$, (ii) for each n, $\alpha(n+1) \sqsubset_{\mathbb{S}} \alpha(n)$ and (iii) for each n there exists p such that $length_{\mathbb{S}}(\alpha(p)) \leqslant_{\mathbb{Q}} \frac{1}{2^n}$. We let $\mathbb{R}$ be the set of real numbers.

Note that we are using a somewhat narrow definition of the notion of a real number, as, in (ii), we require: for each n, $\alpha(n+1) \sqsubset_{\mathbb{S}} \alpha(n)$, rather than: for each n, $\alpha(n+1) \sqsubseteq_{\mathbb{S}} \alpha(n)$.

We introduce binary relations $\#_{\mathbb{R}}$ and $=_{\mathbb{R}}$ on $\mathcal{N}$ such that, for all x, y in $\mathbb{R}$,

$x \#_{\mathbb{R}} y$ (x *is really-apart from* y) if and only if, for some n, $x(n) \#_{\mathbb{S}} y(n)$,
and $x =_{\mathbb{R}} y$ (x *really-coincides with* y) if and only if, for all n, $x(n) \approx_{\mathbb{S}} y(n)$.

The usual functions and operations on $\mathbb{R}$ are defined straightforwardly. We will omit subscripts $\mathbb{Q}$, $\mathbb{R}$, if confusion seems unlikely.

Let X be a subset of $\mathbb{N}$. X is a *partial continuous function from* $\mathbb{R}$ *to* $\mathbb{R}$ if and only if (i) for all m in X there exist s, t in $\mathbb{S}$ such that $m = \langle s, t\rangle$ and (ii) for all s, t, u in $\mathbb{S}$, if $\langle s, t\rangle$ belongs to X and $u \sqsubseteq_{\mathbb{S}} s$, then $\langle u, t\rangle$ belongs to X, and (iii) for all s, t, u

in $\mathbb{S}$, if $\langle s, t\rangle$ belongs to X and $t \sqsubseteq_{\mathbb{S}} u$, then $\langle s, u\rangle$ belongs to X, and (iv) for all s, t and u in $\mathbb{S}$, if both $\langle s, t\rangle$ and $\langle s, u\rangle$ belong to X, then $t \approx_{\mathbb{S}} u$.

Let X be a partial continuous function from $\mathbb{R}$ to $\mathbb{R}$ and let x, y be real numbers. We define: *X maps x onto y*, notation: $X : x \mapsto y$, if and only if, for each n, there exists m such that $\langle x(m), y(n)\rangle$ belongs to X. One may verify that, for all real numbers x, y, u, v, if X maps x onto y and u onto v and $x =_{\mathbb{R}} u$, then $y =_{\mathbb{R}} v$. We let the *domain of* X, notation $dom_{\mathbb{R}}(X)$, be the set of all real numbers x such that, for some real number y, X maps x onto y.

For each γ in $\mathcal{N}$, we call the set of all m such that, for some n, $\gamma(n) = m + 1$, the subset of $\mathbb{N}$ *enumerated by* γ, notation: E_γ. A subset X of $\mathbb{N}$ is called *enumerable* if and only if, for some γ, $X = E_\gamma$.

Let ϕ be an element of $\mathcal{N}$ that enumerates a partial continuous function from $\mathbb{R}$ to $\mathbb{R}$ and let x belong to $dom_{\mathbb{R}}(E_\phi)$. We let $\phi(x)$ be the real number y such that, *firstly*, there exist m, p such that $\phi(p) = \langle x(m), y(0)\rangle + 1$ and $length_{\mathbb{S}}(y(0)) < 1$ and there is no $q < p$ such that, for some k in $\mathbb{N}$, for some r in $\mathbb{S}$, $\phi(q) = \langle x(k), r\rangle + 1$ and $length_{\mathbb{S}}(r) < 1$, and *secondly*, for each n, there exist m, p such that $\phi(p) = \langle x(m), y(n + 1)\rangle + 1$ and $length_{\mathbb{S}}(y(n + 1)) <_{\mathbb{Q}} \frac{1}{2^{n+1}}$ and $y(n + 1) \sqsubset_{\mathbb{S}} y(n)$, and there is no $q < p$ such that, for some k in $\mathbb{N}$, for some r in $\mathbb{S}$, $\phi(q) = \langle x(k), r\rangle + 1$ and $length_{\mathbb{S}}(r) <_{\mathbb{Q}} \frac{1}{2^{n+1}}$ and $r \sqsubset_{\mathbb{S}} y(n)$.
Observe that, for every x in $dom_{\mathbb{R}}(E_\phi)$, E_ϕ maps x onto $\phi(x)$.

Let X be a subset of $\mathbb{N}$. X is a *partial continuous function from $\mathcal{N}$ to $\mathbb{R}$* if and only if (i) for all m in X there exist s, t such that $m = \langle s, t\rangle$ and t belongs to $\mathbb{S}$ and (ii) for all s, t, n, if $\langle s, t\rangle$ belongs to X, then $\langle s * \langle n\rangle, t\rangle$ belongs to X, and (iii) for all s, t, u, if $\langle s, t\rangle$ belongs to X and u belongs to $\mathbb{S}$ and $t \sqsubseteq_{\mathbb{S}} u$, then $\langle s, u\rangle$ belongs to X, and (iv) for all s, t and u, if both $\langle s, t\rangle$ and $\langle s, u\rangle$ belong to X, then $t \approx_{\mathbb{S}} u$.

Let X be a partial continuous function from $\mathcal{N}$ to $\mathbb{R}$ and let α belong to $\mathcal{N}$ and x to $\mathbb{R}$. We define: *X maps α onto x*, notation: $X : \alpha \mapsto x$, if and only if, for each n, there exists m such that $\langle \overline{\alpha}m, x(n)\rangle$ belongs to X. We let the *domain of* X, notation $dom(X)$, be the set of all α in $\mathcal{N}$ such that, for some real number x, X maps α onto x.

Let ϕ be an element of $\mathcal{N}$ that enumerates a partial continuous function from $\mathcal{N}$ to $\mathbb{R}$ and let α belong to $dom(E_\phi)$. We let $\phi|\alpha$ be the real number y such that, *firstly*, there exist m, p such that $\phi(p) = \langle \overline{\alpha}m, y(0)\rangle + 1$ and $length_{\mathbb{S}}(y(0)) < 1$ and there is no $q < p$ such that, for some k in $\mathbb{N}$, for some r in $\mathbb{S}$, $\phi(q) = \langle \overline{\alpha}k, r\rangle + 1$ and $length_{\mathbb{S}}(r) < 1$, and *secondly*, for each n, there exist m, p such that $\phi(p) = \langle \overline{\alpha}m, y(n+1)\rangle + 1$ and $length_{\mathbb{S}}(y(n+1)) <_{\mathbb{Q}} \frac{1}{2^{n+1}}$ and $y(n+1) \sqsubset_{\mathbb{S}} y(n)$, and there is no $q < p$ such that , for some k in $\mathbb{N}$, for some r in $\mathbb{S}$, $\phi(q) = \langle \overline{\alpha}k, r\rangle + 1$ and $length_{\mathbb{S}}(r) <_{\mathbb{S}} \frac{1}{2^{n+1}}$ and $r \sqsubset_{\mathbb{S}} y(n)$.

Observe that, for every α in $dom(E_\phi)$, E_ϕ maps α onto $\phi|\alpha$.

We now define binary relations $<_{\mathbb{S}}$ and $\leqslant_{\mathbb{S}}$ on $\mathbb{N}$ such that, for all p, q, r, s in $\mathbb{Q}$ such that $p <_{\mathbb{Q}} q$ and $r <_{\mathbb{Q}} s$,

$\langle p, q\rangle <_{\mathbb{S}} \langle r, s\rangle$ ($\langle p, q\rangle$ *lies to the left of* $\langle r, s\rangle$) if and only if $q <_{\mathbb{Q}} r$,
and $\langle p, q\rangle \leqslant_{\mathbb{S}} \langle r, s\rangle$ ($\langle p, q\rangle$ *does not lie to the right of* $\langle r, s\rangle$) if and only if $p \leqslant_{\mathbb{Q}} s$.

We also define binary relations $<_{\mathbb{R}}$ and $\leqslant_{\mathbb{R}}$ on $\mathbb{R}$ such that, for all real numbers x, y,

$x <_{\mathbb{R}} y$ if and only if, for some n, $x(n) <_{\mathbb{S}} y(n)$,
and $x \leqslant_{\mathbb{R}} y$ if and only if, for all n, $x(n) \leqslant_{\mathbb{S}} y(n)$.

We will repeatedly make use of the fact that the relation $<_{\mathbb{R}}$ is *co-transitive*, that is, for all real numbers x, y, z, if $x <_{\mathbb{R}} y$, then either $x <_{\mathbb{R}} z$ or $z <_{\mathbb{R}} y$.

It is also important that, for all real numbers x, y, $x \leqslant_{\mathbb{R}} y$ if and only if not $y <_{\mathbb{R}} x$.

We sometimes, if confusion seems unlikely, omit subscripts.

For all u, v, p, q in $\mathbb{Q}$ such that $u <_{\mathbb{Q}} v$ and $p <_{\mathbb{Q}} q$ we define $max_{\mathbb{S}}(\langle u, v\rangle, \langle p, q\rangle)$ in $\mathbb{S}$ by: $max_{\mathbb{S}}(\langle u, v\rangle, \langle p, q\rangle) = \langle max(u, p), max(v, q)\rangle$.

For all real numbers x, y we define a real number $sup(x, y)$ such that, for all n $sup(x, y)(n) = max_{\mathbb{S}}(x(n), y(n))$. One may prove that, for all real numbers x, y, z, $x \leqslant sup(x, y)$ and $y \leqslant sup(x, y)$, and if $x \leqslant z$ and $y \leqslant z$, then $sup(x, y) \leqslant z$.

We let $0_{\mathbb{R}}$ and $1_{\mathbb{R}}$ be real numbers such that, for all n, $0_{\mathbb{R}}(n) = \langle -\frac{1}{2^n}, \frac{1}{2^n}\rangle$ and $1_{\mathbb{R}}(n) = \langle 1 - \frac{1}{2^n}, 1 + \frac{1}{2^n}\rangle$.

We let $[0, 1]$ be the set of all real numbers x such that $0_{\mathbb{R}} \leqslant_{\mathbb{R}} x \leqslant_{\mathbb{R}} 1_{\mathbb{R}}$.

More generally, if y, z are real numbers such that $y \leqslant_{\mathbb{R}} z$, we let $[y, z]$ be the set of all real numbers x such that $y \leqslant_{\mathbb{R}} x \leqslant_{\mathbb{R}} z$.

Let s belong to $\mathbb{N}$. s is called a *binary sequence number* if and only if, for all i, if $i < length(a)$, then $a(i) < 2$. We define a function C from the set of binary sequence numbers to $\mathbb{S}$, as follows, by induction on the length of the sequence numbers:

(i) $C(\langle\ \rangle) = \langle 0, 1\rangle$.
(ii) For all binary sequence numbers s, for all rational numbers p, q, if $C(s) = \langle p, q\rangle$, then $C(s * \langle 0\rangle) = \langle p, \frac{1}{3}p + \frac{2}{3}q\rangle$ and $C(s * \langle 1\rangle) = \langle \frac{2}{3}p + \frac{1}{3}q, q\rangle$.

We let ψ be an element of $\mathcal{N}$ such that E_ψ is a continuous function from $\mathcal{C}$ to $\mathbb{R}$ containing every pair $\langle s, C(s)\rangle$. Note that ψ maps $\mathcal{C}$ onto $[0, 1]$, that is, for every α in $\mathcal{C}$, $\psi|\alpha$ belongs to $[0, 1]$, and, for every x in $[0, 1]$, there exists α in $\mathcal{C}$ such that $\psi|\alpha$ really-coincides with x.

Let X be a partial continuous function from $\mathbb{R}$ to $\mathbb{R}$ and let Y be a subset of $dom_{\mathbb{R}}(X)$. X is *uniformly continuous on* Y if and only if, for each n, there exists m such that for all x, y in Y, if $|x - y| < \frac{1}{2^m}$, then $|\phi(x) - \phi(y)| < \frac{1}{2^n}$. X *positively fails to be uniformly continuous on* Y if and only if there exists n such that, for all m, there are x, y in Y with the property: $|x - y| < \frac{1}{2^m}$ and $|\phi(x) - \phi(y)| \geqslant \frac{1}{2^n}$.

Theorem 1 *The following statements are provable in* BIM*:*
(i) The Weak Fan Theorem implies that every enumerable continuous function from $[0, 1]$ *to* $\mathbb{R}$ *is uniformly continuous on* $[0, 1]$.
(ii) If there exists an enumerable continuous function from $[0, 1]$ *to* $\mathbb{R}$ *that positively fails to be uniformly continuous on* $[0, 1]$*, then Kleene's Alternative holds.*

Proof (i) Let ϕ enumerate a continuous function from $[0, 1]$ to $\mathbb{R}$ and let n be a natural number. Note that, for every α in $\mathcal{C}$, there exist p, i, s such that $\phi(p) = \langle C(\overline{\alpha}i), s\rangle + 1$ and $length_{\mathbb{S}}(s) < \frac{1}{2^n}$. Let B be the set of all binary sequence numbers a such that there exist $p, i, s \leqslant length(a)$ with the property: $\phi(p) = \langle C(\overline{a}i), s\rangle + 1$ and $length_{\mathbb{S}}(s) < \frac{1}{2^n}$. Note that B is a decidable subset of $\mathbb{N}$ and a bar in $\mathcal{C}$. Using the Weak Fan Theorem, find m such that, for every α in $\mathcal{C}$, there exists $j \leqslant m$ such that $\overline{\alpha}j$ belongs to B, and, therefore, $\overline{\alpha}m$ belongs to B. Note that, for all x, y in $[0, 1]$, if $|x - y| < \frac{1}{3^m}$, then there exist α, β in $\mathcal{C}$ such that $\overline{\alpha}m = \overline{\beta}m$ and $\psi|\alpha =_{\mathbb{R}} x$ and $\psi|\beta =_{\mathbb{R}} y$, and, therefore, $|\phi(x) - \phi(y)| < \frac{1}{2^n}$.
(ii) Let ϕ enumerate a continuous function from $[0, 1]$ to $\mathbb{R}$ that positively fails to be uniformly continuous on $[0, 1]$. Find n such that for all m, there are x, y in Y with the property: $|x - y| < \frac{1}{2^m}$ and $|\phi(x) - \phi(y)| \geqslant \frac{1}{2^n}$. Let B be the set of all binary sequence numbers a such that, for some $p, i, s \leqslant length(a)$, $\phi(p) = \langle C(\overline{a}i), s\rangle + 1$ and $length_{\mathbb{S}}(s) < \frac{1}{2^n}$. Note that B is a decidable subset of $\mathbb{N}$ and a bar in $\mathcal{C}$. Also note that every finite subset of B positively fails to be a bar in $\mathcal{C}$. $\square$

In [13], it is shown that the statements converse to Theorem 1(i) and (ii) are also provable in BIM.

In Section 5, we will indicate how to introduce partial continuous functions from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$ or to $\mathbb{R} \times \mathbb{R}$. The Weak Fan Theorem also implies that every enumerable continuous function from $[0, 1] \times [0, 1]$ to $\mathbb{R}$ or from $[0, 1] \times [0, 1]$ to $\mathbb{R} \times \mathbb{R}$ is uniformly continuous, and if a function of one of these kinds positively fails to be uniformly continuous on $[0, 1] \times [0, 1]$, then Kleene's Alternative holds.

4 The One-Dimensional Case

The one-dimensional *Fixed-Point Theorem*, in its usual formulation, is constructively false:

> *For every ϕ, if ϕ enumerates a continuous function from* $[0, 1]$ *to* $[0, 1]$*, then there exists x in* $[0, 1]$ *such that* $\phi(x) = x$.

This theorem is easily seen to be equivalent to the *Intermediate Value Theorem*:

> *For every ϕ, if ϕ enumerates a continuous function from* $[0, 1]$ *to* $\mathbb{R}$ *such that* $\phi(0) = 0$ *and* $\phi(1) = 1$*, then there exists x in* $[0, 1]$ *such that* $\phi(x) = \frac{1}{2}$.

We will construct a *counterexample in Brouwer's style* to the one-dimensional fixed-point theorem.

We first define a special real number ρ that *vacillates around* 0, as follows.

Let d be the decimal expansion of π, that is, d belongs to $\mathcal{N}$, and for each j, $d(j) < 10$, and $\pi = 3 + \sum_{i \in \mathbb{N}} \frac{d(i)}{10^{i+1}}$. For each n, if there is no $i \leqslant n$ such that for all $j < 100$, $d(i + j) = 9$, then $\rho(n) = \langle -\frac{1}{2^n}, \frac{1}{2^n}\rangle$, and, if there is such i, then $\rho(n) = \langle (-1)^k \frac{1}{2^{k+1}} - \frac{1}{2^{n+2}}, (-1)^k \frac{1}{2^{k+1}} + \frac{1}{2^{n+2}}\rangle$, where k is the least such i. Note that, if $\rho \leqslant 0$, then, for each k, if k is the least i such that for all $j < 100$, $d(i + j) = 9$, then k is odd, and, if $\rho \geqslant 0$, then, for each k, if k is the least i such that for all $j < 100$, $d(i + j) = 9$, then k is even, so we have no proof that $\rho \leqslant 0$ and we have no proof that $\rho \geqslant 0$.

The statements "$\rho \leqslant 0$" and "$\rho \geqslant 0$" are equivalent to the statements "$sup(\rho, 0) = 0$" and "$sup(\rho, 0) = \rho$", respectively.

Finally, let ϕ enumerate a continuous function from $[0, 1]$ to $[0, 1]$ such that $\phi(0) = \frac{1}{3}$, $\phi(\frac{1}{3}) = \frac{1}{3} + \rho$, $\phi(\frac{2}{3}) = \frac{2}{3} + \rho$ and $\phi(1) = \frac{2}{3}$ and ϕ is linear on $[0, \frac{1}{3}]$, on $[\frac{1}{3}, \frac{2}{3}]$ and on $[\frac{2}{3}, 1]$.

Note that, if $\rho > 0$, then, for all y in $[0, \frac{2}{3}]$, $\phi(y) > y$ and, if $\rho < 0$, then, for all y in $[\frac{1}{3}, 1]$, $\phi(y) < y$. Suppose we find x in $[0, 1]$ such that $\phi(x) = x$. If $x > \frac{1}{3}$, then $\rho \geqslant 0$, and if $y < \frac{2}{3}$, then $\rho \leqslant 0$. So we are unable to find such x.

Now that the one-dimensional Fixed-Point Theorem has been shown to be false, we may derive some consolation from the fact that, using the method of iterated bisection, we can prove an approximate version:

Theorem 2 *The following statement is provable in* BIM*:*
For every ϕ, if ϕ enumerates a continuous function from $[0, 1]$ *to* $[0, 1]$*, then, for every positive rational number e there exists x in* $[0, 1]$ *such that* $|\phi(x) - x| < e$.

Proof Suppose that ϕ belongs to $\mathcal{N}$ and enumerates a continuous function from $[0, 1]$ to $[0, 1]$. Let e be a positive rational number. We construct two infinite sequences α, β of rational numbers, as follows, by induction.

(i) $\alpha(0) = 0$ and $\beta(0) = 1$. Observe that $\phi(\alpha(0)) > \alpha(0) - \frac{1}{6}e$ and $\phi(\beta(0)) < \beta(0) + \frac{1}{6}e$.

(ii) Suppose that n belongs to $\mathbb{N}$ and that $\alpha(n)$ and $\beta(n)$ have been defined already. We consider $c := \frac{\alpha(n)+\beta(n)}{2}$. We now determine the first k such that, for some u, v, p, q in $\mathbb{Q}$, $\phi(k) = \langle\langle u, v\rangle, \langle p, q\rangle\rangle + 1$ and $u < c < v$ and $q - p < \frac{1}{3}e$. Let l be this number. Find u, v, p, q in $\mathbb{Q}$ such that $\phi(l) = \langle\langle u, v\rangle, \langle p, q\rangle\rangle + 1$. We distinguish two cases:
Case (1): $q < c + \frac{1}{6}e$. We define: $\alpha(n + 1) = \alpha(n)$ and $\beta(n + 1) = c$.
Case (2): $q \geqslant c + \frac{1}{6}e$. We may conclude that $p > c - \frac{1}{6}e$. We define: $\alpha(n+1) = c$ and $\beta(n + 1) = \beta(n)$.

Observe that, for each n, $\phi(\alpha(n)) > \alpha(n) - \frac{1}{6}e$ and $\phi(\beta(n)) < \beta(n) + \frac{1}{6}e$ and $\beta(n) - \alpha(n) = \frac{1}{2^n}$.

Let x be a real number such that, for each n, $\langle\alpha(n), \beta(n)\rangle \sqsubset_{\mathbb{S}} x(n)$. Note that, for each n, x belongs to $[\alpha(n), \beta(n)]$.

Find the least k such that, for some u, v, p, q in $\mathbb{Q}$, $\phi(k) = \langle\langle u, v\rangle, \langle p, q\rangle\rangle + 1$, and, for some $i < k$, $x(i) \sqsubset_{\mathbb{S}} \langle u, v\rangle$ and $\beta(i) - \alpha(i) < \frac{1}{3}e$ and $q - p < \frac{1}{6}e$. Let l be this number. Find u, v, p, q in $\mathbb{Q}$ such that $\phi(l) = \langle\langle u, v\rangle, \langle p, q\rangle\rangle + 1$. Determine i such that $\beta(i) - \alpha(i) < \frac{1}{3}e$ and $x(i) \sqsubset_{\mathbb{S}} \langle u, v\rangle$. Observe that $q - p < \frac{1}{6}e$ and that $\langle p, q\rangle$ will contain a rational number greater than $\alpha(i) - \frac{1}{6}e$ and also a rational number smaller than $\beta(i) + \frac{1}{6}e$. It follows that $p > \alpha(i) - \frac{1}{3}e$ and $q < \beta(i) + \frac{1}{3}e$ and that $\langle p, q\rangle \sqsubseteq_{\mathbb{S}} \langle\alpha(i) - \frac{1}{3}e, \beta(i) + \frac{1}{3}e\rangle$. Therefore, both x and $\phi(x)$ belong to $[\alpha(i) - \frac{1}{3}e, \beta(i) + \frac{1}{3}e]$. Note that $(\beta(i) + \frac{1}{3}e) - (\alpha(i) - \frac{1}{3}e) < e$ and, therefore, $|x - \phi(x)| < e$. □

5 A Simple Observation on the Two-Dimensional Case

Let α belong to $\mathcal{N}$ and m to $\mathbb{N}$. We let α^m be the element β of $\mathcal{N}$ such that, for every n, $\beta(n) = \alpha(J(m, n))$.

Let X, Y be subsets of $\mathcal{N}$. We let $X \times Y$ be the set of all α in $\mathcal{N}$ such that α^0 belongs to X and α^1 belongs to Y. We let X^2 be $X \times X$.

For all α, β in $\mathcal{N}$ we let $\langle \alpha, \beta \rangle$ be the element γ of $\mathcal{N}$ such that $\gamma^0 = \alpha$ and $\gamma^1 = \beta$ and, for each $i > 1$, for all n, $\gamma^i(n) = 0$. We sometimes feel free to write: "(α, β)" rather than "$\langle \alpha, \beta \rangle$".

Let X be a subset of $\mathbb{N}$. X *is a partial continuous function from* $\mathbb{R}^2$ *to* $\mathbb{R}$ if and only if (i) for each n in X there exist r, s, t in $\mathbb{S}$ such that $n = \langle \langle r, s \rangle, t \rangle$, and (ii) for all r, s, t, v, w in $\mathbb{S}$, if $\langle \langle r, s \rangle, t \rangle$ belongs to X and $v \sqsubseteq_{\mathbb{S}} r$ and $w \sqsubseteq_{\mathbb{S}} s$, then $\langle \langle v, w \rangle, t \rangle$ belongs to X, and (iii) for all r, s, t, v in $\mathbb{S}$, if $\langle \langle r, s \rangle, t \rangle$ belongs to X, and $t \sqsubseteq_{\mathbb{S}} v$, then $\langle \langle r, s \rangle, v \rangle$ belongs to X, and (iv) for all r, s, t, u, v, w in $\mathbb{S}$, if both $\langle \langle r, s \rangle, t \rangle$ and $\langle \langle r, s \rangle, v \rangle$ belong to X, then $t \approx_{\mathbb{S}} v$.

Let X be a partial continuous function from $\mathbb{R}^2$ to $\mathbb{R}$ and let x, y, z be real numbers. X *maps* $\langle x, y \rangle$ *onto* z, notation: $X : \langle x, y \rangle \mapsto z$, if and only if for each n there exists m such that $\langle \langle x(m), y(m) \rangle, z(n) \rangle$ belongs to X.
The domain and the range of X, notation: $dom_{\mathbb{R}^2}(X)$ and $ran_{\mathbb{R}}(X)$, are defined straightforwardly.

Let ϕ in $\mathcal{N}$ enumerate a partial continuous function from $\mathbb{R}^2$ to $\mathbb{R}$. We denote the domain and the range of E_ϕ by $dom_{\mathbb{R}^2}(\phi)$ and $ran_{\mathbb{R}}(\phi)$, respectively. For all real numbers x, y, if $\langle x, y \rangle$ belongs to $dom_{\mathbb{R}^2}(\phi)$, we let $\phi(x, y)$ be the number z such that, for each n, there exists p such that, for some m, $\phi(p) = \langle \langle x(m), y(m) \rangle, z(n) \rangle + 1$, and $length_{\mathbb{S}}(z(n)) < \frac{1}{2^n}$ and, if $n > 0$, then $z(n) \sqsubseteq_{\mathbb{S}} z(n-1)$, *and* there do not exist q, t such that $q < p$ and, for some m, $\phi(q) = \langle \langle x(m), y(m) \rangle, t \rangle + 1$, and $length_{\mathbb{S}}(t) < \frac{1}{2^n}$ and, if $n > 0$, then $t \sqsubseteq_{\mathbb{S}} z(n-1)$.
Observe that, for every $\langle x, y \rangle$ in $dom_{\mathbb{R}^2}(\phi)$, E_ϕ maps $\langle x, y \rangle$ onto $\phi(x, y)$.

Suppose that ϕ belongs to $\mathcal{N}$. ϕ *enumerates a partial continuous function from* $\mathbb{R}^2$ *to* $\mathbb{R}^2$ if and only if both ϕ^0 and ϕ^1 enumerate a partial continuous function from $\mathbb{R}^2$ to $\mathbb{R}$.

Suppose that ϕ enumerates a partial continuous function from $\mathbb{R}^2$ to $\mathbb{R}^2$. We let the *domain of* ϕ, notation: $dom_{\mathbb{R}^2}(\phi)$, be the set of all pairs of real numbers that belong to both $dom_{\mathbb{R}^2}(\phi^0)$ and $dom_{\mathbb{R}^2}(\phi^1)$. For all pairs $\langle x, y \rangle$ in $dom_{\mathbb{R}^2}(\phi)$ we define $\phi(x, y) := \langle \phi^0(x, y), \phi^1(x, y) \rangle$.

We will use $|\ |$ to denote the usual euclidean distance on $\mathbb{R}^2$.

From now on, we denote the unit square $[0, 1] \times [0, 1]$ by U. We let ∂U be the *border of* U, that is, the set of all points p in $\mathbb{R}^2$ with the property that, for every positive rational number e, there exists $\langle x, y \rangle$ in U such that $|p - \langle x, y \rangle| < e$ and either $x = 0$ or $x = 1$ or $y = 0$ or $y = 1$. A point (x, y) *is an interior point of* U if and only if both $0 < x < 1$ and $0 < y < 1$.

Theorem 3 *The following statement is provable in* BIM*:*
For every ϕ enumerating a continuous function from U to U, for every positive rational number e, if $e > \frac{1}{2}$, then there exists p in U such that $|\phi(p) - p| < e$.

Proof Let ϕ enumerate a continuous function from U to U. We let ψ enumerate a continuous function from $[0, 1]$ to $[0, 1]$, such that, for every y in $[0, 1]$, $\psi(y) = \phi^1(\frac{1}{2}, y)$.

Let e be a rational number and assume $e > \frac{1}{2}$. Find a rational number b such that $b^2 < e^2 - \frac{1}{4}$. Using Theorem 2, find y in $[0, 1]$ such that $|\psi(y) - y| < b$. Observe that the point $p = \langle \frac{1}{2}, y\rangle$ satisfies our purposes, as $|\phi(\frac{1}{2}, y) - \langle \frac{1}{2}, y\rangle| = \sqrt{(\phi^0(\frac{1}{2}, y) - \frac{1}{2})^2 + (\phi^1(\frac{1}{2}, y) - y)^2} \leqslant \sqrt{\frac{1}{4} + b^2} < e$. □

6 Preparing for the Main Results

In this section, we prove two results that we want to use in the next section.

Let X be a finite set of (non-degenerate) rational segments. We let $\overline{X}$, the *closure of* X, be the set of all real numbers x such that, for each n, there exists s in X with the property: $s \approx_{\mathbb{S}} x(n)$. We call X *almost-disjoint* if and only if, for all p, q, r, s in $\mathbb{Q}$, if both $p <_{\mathbb{Q}} q$ and $r <_{\mathbb{Q}} s$, and both $\langle p, q\rangle$ and $\langle r, s\rangle$ belong to X, and $\langle p, q\rangle \neq \langle r, s\rangle$, then either $q \leqslant_{\mathbb{Q}} r$ or $s \leqslant_{\mathbb{Q}} p$. Speaking geometrically, X is almost-disjoint if different elements of X have at most endpoints in common.

Let ϕ, ψ enumerate partial continuous functions from $\mathbb{R}^2$ to $\mathbb{R}^2$. We say that *(the function enumerated by)* ψ *extends (the function enumerated by)* ϕ if and only if every element of $dom_{\mathbb{R}^2}(\phi)$ really-coincides with an element of $dom_{\mathbb{R}^2}(\psi)$, and, for every $\langle u, v\rangle$ in $dom_{\mathbb{R}^2}(\phi)$, for every $\langle x, y\rangle$ in $dom_{\mathbb{R}^2}(\psi)$, if $u =_{\mathbb{R}} x$ and $v =_{\mathbb{R}} y$, then $\phi^0(u, v) =_{\mathbb{R}} \psi^0(x, y)$ and $\phi^1(u, v) =_{\mathbb{R}} \psi^1(x, y)$.

Lemma 4 *(i) For all rational segments* s, t, *if* $s \sqsubseteq_{\mathbb{S}} \langle 0, 1\rangle$ *and* $t \sqsubseteq_{\mathbb{S}} \langle 0, 1\rangle$ *and at least one of* s, t *does not coincide with* $\langle 0, 1\rangle$*, then there exists* ϕ *enumerating a continuous function from* $\overline{\{s\}} \times \overline{\{t\}}$ *to* ∂U *such that, for every* p *in* $\overline{\{s\}} \times \overline{\{t\}}$*, if* p *belongs to* ∂U*, then* $\phi(p) = p$.
(ii) Let X *be an almost-disjoint finite subset of* $\mathbb{S}$ *such that* $\overline{X}$ *is a proper subset of* $[0, 1]$. *Let* ϕ *enumerate a continuous function from* $\overline{X} \times \overline{X}$ *to* ∂U *such that for every* p *in* $\overline{X} \times \overline{X}$*, if* p *belongs to* ∂U*, then* $\phi(p) = p$. *Let* s *be an element of* $\mathbb{S}$ *such that* $X \cup \{s\}$ *is almost-disjoint and* $\overline{X \cup \{s\}}$ *is a proper subset of* $[0, 1]$.
There exists ψ *in* $\mathcal{N}$ *such that* ψ *enumerates a continuous function from* $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ *to* ∂U *and* ψ *extends* ϕ *and for every* p *in* $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$*, if* p *belongs to* ∂U*, then* $\phi(p) = p$.

Proof (i) We leave the proof to the reader.
(ii) Let X, ϕ and s satisfy the requirements. One has to distinguish four cases.

> *Case (i):* There is no t in X such that $s \approx_{\mathbb{S}} t$.
> *Case (ii):* There is exactly one t in X such that $s \approx_{\mathbb{S}} t$ and there are rational numbers p, q such that $0 < p < q < 1$ such that $s = \langle p, q\rangle$, that is neither 0 nor 1 belongs to $\overline{\{s\}}$.
> *Case (iii):* There is exactly one t in X such that $s \approx_{\mathbb{S}} t$ and there is a rational number p such that either $s = \langle 0, p\rangle$ or $s = \langle p, 1\rangle$.
> *Case (iv):* There are exactly two elements t of X such that $s \approx_{\mathbb{S}} t$.

Note that there exist k, l such that $\overline{X} \times \overline{X}$ consists of k^2 mutually disjoint rectangles and $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ consists of l^2 mutually disjoint rectangles.

In Case (i) $l = k + 1$. The rectangles making up $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ are the rectangles contained in $\overline{X} \times \overline{X}$ and $2k + 1$ new rectangles, each of them disjoint from every rectangle in $\overline{X} \times \overline{X}$. Using (i), we find an enumerable extension of the function enumerated by ϕ that satisfies the requirements.

In case (ii) $l = k$. The number of rectangles in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ is the same as the number of rectangles in $\overline{X} \times \overline{X}$. However, there are $2k - 1$ rectangles in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ that are proper extensions of rectangles in $\overline{X} \times \overline{X}$. The extension may take one of the three forms sketched in Fig. 1. Note that in each case we easily construct an enumerable continuous function from the extended rectangle to the unextended rectangle that maps every point in the unextended rectangle onto itself. Combining these functions we find ρ enumerating a *retraction* of $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ onto $\overline{X} \times \overline{X}$, that is, a continuous mapping of $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ onto $\overline{X} \times \overline{X}$ leaving every point of $\overline{X} \times \overline{X}$ invariant, and we let ψ be an element of $\mathcal{N}$ enumerating a continuous function from $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ to U such that, for each point p in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$, $\psi(p) = (\phi(\rho(p)))$.

In case (iii) $l = k$ but either 0 or 1 is an endpoint of s and all the extended rectangles have sides on the border ∂U of the unit square U.

Let us consider, for instance, in Fig. 2, rectangles $ABCD$ and $EFGH$ that are extended to rectangles $PBCQ$ and $RFST$, respectively, where the sides PQ, RT and TS are on the border ∂U of U.

We first make an enumerable function from the segment QD to ∂U that maps Q onto itself and D onto $\phi(D)$. Using this function, we find an enumerable partial continuous function mapping defined on $\overline{PQ \cup QD \cup ABCD}$ to ∂U, mapping every

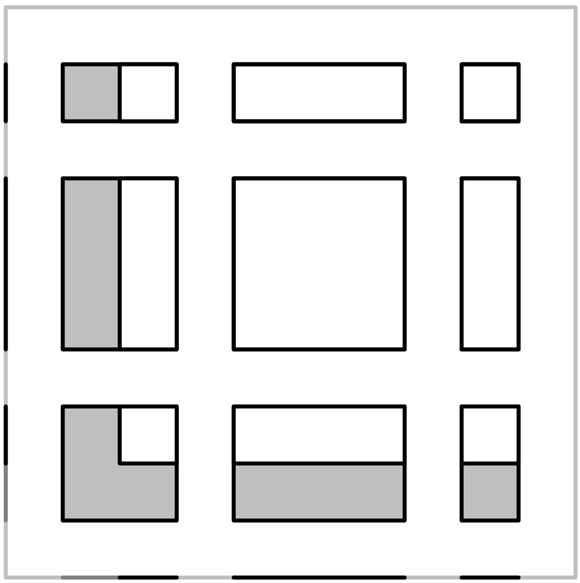

Fig.1

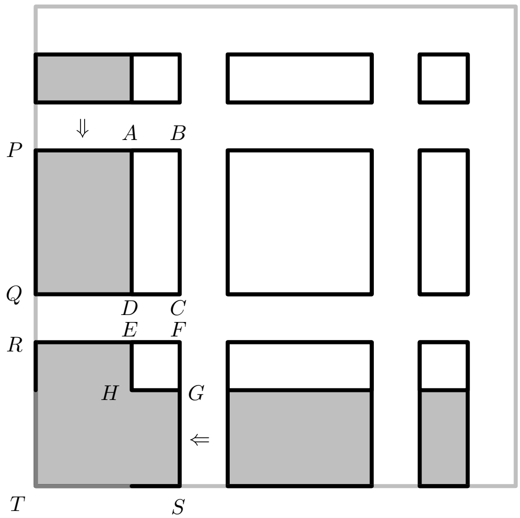

Fig.2

point in PQ onto itself and extending the function enumerated by ϕ on $ABCD$. We now make an enumerable retraction of $PBCQ$ onto $\overline{PQ \cup QD \cup ABCD}$, by 'pushing' PA and the whole rectangle $PAQD$ onto $\overline{PQ \cup QD \cup DA}$ and leaving $\overline{PQ \cup QD \cup ABCD}$ invariant. Taking the composition of the function mentioned earlier and this retraction, we obtain an enumerable function of the square $PBCQ$ to ∂U extending the function enumerated by ϕ on $ABCD$ and mapping every point in PQ onto itself.

The square $EFGH$ and its extension $RFST$ are treated similarly. We first make an enumerable function from the segment RE to ∂U that maps R onto itself and E onto $\phi(E)$. Using this function, we find an enumerable partial continuous function mapping defined on $\overline{ST \cup TR \cup RE \cup EFGH}$ to ∂U, mapping every point in $ST \cup TR$ onto itself and extending the function enumerated by ϕ on $EFGH$. We now make an enumerable retraction of $RFST$ onto $\overline{ST \cup TR \cup RE \cup EFGH}$, by 'pushing' GS and the whole hexagon $GSTREH$ onto $\overline{ST \cup TR \cup RE \cup EH \cup HG}$ and leaving $\overline{ST \cup TR \cup RE \cup EFGH}$ invariant. Taking the composition of the function mentioned earlier and this retraction, we obtain an enumerable function of the square $RFST$ to ∂U extending the function enumerated by ϕ on $EFGH$ and mapping every point in $\overline{ST \cup TR}$ onto itself.

Treating all rectangles in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ in this way, we obtain ψ in $\mathcal{N}$, enumerating a continuous function from $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ to ∂U extending the function enumerated by ϕ and such that, for every p in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$, if p belongs to ∂U, then $\psi(p) = p$.

In case (iv) $l = k - 1$. We see that 4 rectangles in $\overline{X} \times \overline{X}$ combine to give one rectangle in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$, and $2k - 2$ pairs of rectangles in $\overline{X} \times \overline{X}$ combine

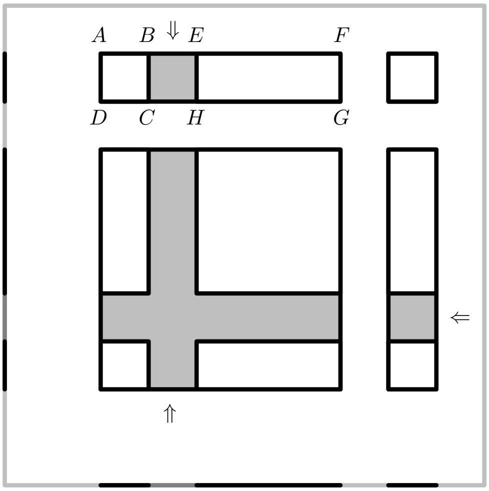

Fig.3

to give $2k-2$ rectangles in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$. We have to show how to extend the function enumerated by ϕ to an enumerable function on each new rectangle.

Considering Fig. 3, we only study the case of the old rectangles $ABCD$ and $EFGH$, together giving rise to the new rectangle $AFGD$, leaving the other cases to the reader. We first make an enumerable function from the segment CH to ∂U that maps C onto $\phi(C)$ and H onto $\phi(H)$. Using this function, we find an enumerable partial continuous function mapping defined on $\overline{ABCD \cup CH \cup EFGH}$ to ∂U, extending the function enumerated by ϕ on $ABCD \cup EFGH$. We now make an enumerable retraction of the new rectangle $AFGD$ onto $\overline{ABCD \cup CH \cup EFGH}$, by 'pushing' BE and the whole rectangle $BEHC$ onto $\overline{BC \cup CH \cup HE}$ and leaving $\overline{ABCD \cup CH \cup EFGH}$ invariant. Taking the composition of the function mentioned earlier and this retraction, we obtain an enumerable function of the square $AFGD$ to ∂U extending the function enumerated by ϕ on $ABCD \cup EFGH$.
Treating all rectangles in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ in this way, we again obtain ψ in $\mathcal{N}$ enumerating a continuous function from $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$ to ∂U extending the function enumerated by ϕ and such that, for every p in $\overline{X \cup \{s\}} \times \overline{X \cup \{s\}}$, if p belongs to ∂U, then $\psi(p) = p$. □

Rational numbers of the form $\frac{m}{2^n}$, where m is an integer and n a natural number, will be called *dyadic rationals*.

There do exist real numbers for which we are unable to decide if they coincide with a dyadic rational or are apart from every dyadic rational. An example of such a number is $\rho\sqrt{2}$, where ρ is the real number vacillating around 0 that we constructed in Section 4. Therefore, the set of all real numbers x in $[0, 1]$ that either really-coincide with a dyadic rational or are really-apart from every dyadic rational, is a proper subset of $[0, 1]$.

Lemma 5 *(i) Let Y be an enumerable partial continuous function from $\mathbb{R}$ to $\mathbb{R}$ such that every real number in* $[0, 1]$ *that either really-coincides with a dyadic rational or is really-apart from every dyadic rational belongs to $dom_{\mathbb{R}}(Y)$. Then* $[0, 1]$ *is a subset of $dom_{\mathbb{R}}(Y)$.*
(ii) Let Y be an enumerable partial continuous function from $\mathbb{R}^2$ to $\mathbb{R}$ such that, for every p in U, if for both $i < 2$, p^i either really-coincides with a dyadic rational or is really-apart from every dyadic rational, then p belongs to $dom_{\mathbb{R}^2}(Y)$. Then U is a subset of $dom_{\mathbb{R}^2}(Y)$.

Proof (i). Let Y satisfy the requirements. We will prove that, for every x in $[0, 1]$, for every n in $\mathbb{N}$, there exist i in $\mathbb{N}$ and t in $\mathbb{S}$ such that $\langle x(i), t\rangle$ belongs to Y and $length_{\mathbb{S}}(t) \leqslant \frac{1}{2^n}$.

Let n belong to $\mathbb{N}$. Using the Minimal Axiom of Countable Choice, we determine a function ε from $\mathbb{N}$ to $\mathbb{N}$ such that for every dyadic rational q in $[0, 1]$ there exists t in $\mathbb{S}$ such that $\langle\langle max(0, q - \varepsilon(q)), min(q + \varepsilon(q), 1)\rangle, t\rangle$ belongs to Y and $length_{\mathbb{S}}(t) \leqslant \frac{1}{2^n}$.

Now let x belong to $[0, 1]$. Using once more the Minimal Axiom of Countable Choice, we determine α in $\mathcal{C}$ such that, for each n, if $\alpha(n) = 0$ and n is a dyadic rational in $[0, 1]$, then $|x - n| > \frac{1}{2}\varepsilon(n)$, and, if $\alpha(n) = 1$, then n is a dyadic rational and $|x - n| < \varepsilon(n)$. We then define a real number x^* in $[0, 1]$, as follows. For each n, if $\overline{\alpha}(n) = \overline{0}(n)$, then $x^*(n) = x(n)$, and if $\overline{\alpha}(n) \neq \overline{0}(n)$, then n will be positive, and $x^*(n)$ will be the first element $s = \langle p, q\rangle$ of $\mathbb{S}$ such that $s \sqsubset_{\mathbb{S}} x^*(n-1)$ and $length_{\mathbb{S}}(s) \leqslant \frac{1}{2^n}$ and $p <_{\mathbb{Q}} q$ and for every $r < n$, if r is a dyadic rational from $[0, 1]$, then either $r <_{\mathbb{Q}} p$ or $q <_{\mathbb{Q}} r$. Note that x^* is really-apart from every dyadic rational, and that, if x is really-apart from x^*, then there exists a dyadic rational q in $[0, 1]$ such that $|x - q| < \varepsilon(q)$.

We determine p in $\mathbb{N}$ and t in $\mathbb{S}$ such that $\langle x^*(p), t\rangle$ belongs to Y and $length_{\mathbb{S}}(t) \leqslant \frac{1}{2^n}$. Observe that, as x^* is a real number in the sense of our rather narrow definition, $x^*(p+1) \sqsubset_{\mathbb{S}} x^*(p)$. Find n such that for every s in $\mathbb{S}$, if $length_{\mathbb{S}}(s) < \frac{1}{2^n}$ and $s \approx_{\mathbb{S}} x^*(p+1)$, then $s \sqsubset x^*(p)$. Find q such that $length_{\mathbb{S}}(x(q)) < \frac{1}{2^n}$ and distinguish two cases:

Case (i): $x(q) \approx_{\mathbb{S}} x^*(p+1)$ and, therefore, $x(q) \sqsubset_{\mathbb{S}} x^*(p)$, and $\langle x(q), t\rangle$ belongs to Y.

Case(ii): Not: $x(q) \approx_{\mathbb{S}} x^*(p+1)$, and, therefore, x is really-apart from x^* and there exists a dyadic rational q in $[0, 1]$ such that $|x - q| < \varepsilon(q)$. We may find i such that $x(i) \sqsubset_{\mathbb{S}} \langle max(0, q - \varepsilon(q)), min(q + \varepsilon(q), 1)\rangle$ and s in $\mathbb{S}$ such that $\langle x(i), s\rangle$ belongs to Y and $length_{\mathbb{S}}(s) \leqslant \frac{1}{2^n}$.

Observe that in both cases there exist i in $\mathbb{N}$ and t in $\mathbb{S}$ such that $\langle x(i), t\rangle$ belongs to Y and $length_{\mathbb{S}}(t) \leqslant \frac{1}{2^n}$.

We may conclude that every element x of $[0, 1]$ belongs to $dom_{\mathbb{R}}(Y)$.
(ii) The proof is similar to the proof of (i). □

7 The Main Results

Theorem 6 *(A constructive version of Sperner's Lemma)*
For every positive rational e, for every ϕ, if ϕ enumerates a uniformly continuous function from U to U, then there exists p in U such that $|\phi(p) - p| < e$.

Proof Let ϕ enumerate a uniformly continuous function from U to U, and let e be a positive rational number.

First observe that, for every point (x, y) in U we can make the following two decisions: (1) *either* $\phi^0(x, y) > x - \frac{1}{6}e$ *or* $\phi^0(x, y) < x + \frac{1}{6}e$ and (2) *either* $\phi^1(x, y) > y - \frac{1}{6}e$ *or* $\phi^1(x, y) < y + \frac{1}{6}e$.

We determine n such that $\frac{1}{2^n} < \frac{1}{6}e$ and, for all p, q in U, if $|p - q| < \frac{1}{2^n}$, then $|\phi(p) - \phi(q)| < \frac{1}{6}e$. We now consider all points in U of the form $(\frac{m}{2^{n+1}}, \frac{p}{2^{n+1}})$, where $m, p \leqslant 2^{n+1}$. We then label these points, attaching to each point p a label $l(p)$ from the set $\{0, 1, 2\}$, taking care that, for all (x, y) in U,

(1) if $l(x, y) = 0$, then $\phi^0(x, y) > x - \frac{1}{6}e$, that is, (x, y) is *not too far westward moving*, and,

(2) if $l(x, y) = 1$, then both $\phi^0(x, y) < x + \frac{1}{6}e$ and $\phi^1(x, y) > y - \frac{1}{6}e$, that is, (x, y) is both *not too far eastward moving* and *not too far southward moving*, and,

(3) if $l(x, y) = 2$, then both $\phi^0(x, y) < x + \frac{1}{6}e$ and $\phi^1(x, y) < y + \frac{1}{6}e$, that is, (x, y) is both *not too far eastward moving* and *not too far northward moving*, and,

(4) if (x, y) belongs to the *westside* $\{0\} \times [0, 1]$ of U, then $l(x, y) = 0$, and, if (x, y) belongs to the *northside* $[0, 1] \times \{1\}$ of U, then $l(x, y) = 0$ or $l(x, y) = 2$, and, if (x, y) belongs to the *southside* $[0, 1] \times \{0\}$ of U, then $l(x, y) = 0$ or $l(x, y) = 1$, and, if (x, y) belongs to the *eastside* $\{1\} \times [0, 1]$ of U, then $l(x, y) = 1$ or $l(x, y) = 2$.

Note that $l(1, 1) = 2$ and $l(1, 0) = 1$.

We investigate triples of one of the two forms
$\big((\frac{m}{2^{n+1}}, \frac{p}{2^{n+1}}), (\frac{m+1}{2^{n+1}}, \frac{p}{2^{n+1}}), (\frac{m+1}{2^{n+1}}, \frac{p+1}{2^{n+1}})\big)$ or
$\big((\frac{m}{2^{n+1}}, \frac{p}{2^{n+1}})(\frac{m}{2^{n+1}}, \frac{p+1}{2^{n+1}}), (\frac{m+1}{2^{n+1}}, \frac{p+1}{2^{n+1}})\big)$, where $m, p < 2^{n+1}$, and call them *elementary triangles*. Consider all sides of elementary triangles lying on ∂U and observe that, if the endpoints of such a side, say pq, receives the two labels 1 and 2, then the two points p, q both lie on the eastside $\{1\} \times [0, 1]$ of U. It follows that the number of sides of elementary triangles lying on ∂U whose endpoints receive the two labels 1 and 2 is odd.

Let C be the set of elementary triangles whose vertices receive the three labels 0, 1 and 2. Suppose that C has an even number of elements. Note that every elementary triangle not in C has either 0 or 2 sides whose endpoints receive the two labels 1 and 2. All elementary triangles together thus have an even number of sides whose endpoints receive the two labels 1 and 2. This number must be odd, however, as there is an odd number of such sides lying on ∂U, and every side lying inside U is counted twice. It follows that C has an odd number of elements, and, therefore, at least one element.

Find an elementary triangle pqr such that its vertices p, q, r receive the labels 0, 1, 2. Take any point (x, y) contained in this triangle. Note that one of the points p, q, r, say $z = (z_0, z_1)$ is not too far westward moving, that is: $\phi^0(z) > z_0 - \frac{1}{6}e$. Note that $|(x, y) - (z_0, z_1)| < \frac{1}{2^n}$. Therefore, both $|x - z_0| < \frac{1}{6}e$ and $|\phi^0(z_0, z_1) - \phi^0(x, y)| < \frac{1}{6}e$. It follows that $\phi^0(x, y) > x - \frac{1}{2}e$.

In a similar way one proves $\phi^0(x, y) < x + \frac{1}{2}e$ and $y - \frac{1}{2}e < \phi^1(x, y) < y + \frac{1}{2}e$. We thus see that $|(x, y) - \phi(x, y)| < e$. □

Theorem 7 *The following statements are equivalent in* BIM.
(i) The Weak Fan Theorem.
(ii) For every positive rational e, for every ϕ, if ϕ enumerates a continuous function from U to U, then there exists p in U such that $|\phi(p) - p| < e$.
(iii) There exists a positive rational number e such that $e < \frac{1}{2}$ and, for every ϕ, if ϕ enumerates a continuous function from U to U, then there exists p in U such that $|\phi(p) - p| < e$.
(iv) For every ϕ, if ϕ enumerates a continuous function from U to U with the property that, for every point p in ∂U, $\phi(p) = p$, then there is a point p in U such that $\phi(p)$ is an interior point of U.

Proof (i) ⇒ (ii). Let e be a positive rational number and let ϕ enumerate a continuous function from U to U. The Weak Fan Theorem implies that the function enumerated by ϕ is uniformly continuous, so we may apply Theorem 6.
(ii) ⇒ (iii). Obvious.
(iii) ⇒ (iv). Let e be a positive rational number such that $e < \frac{1}{2}$ and, for every ϕ, if ϕ enumerates a continuous function from U to U, then there exists p in U such that $|\phi(p) - p| < e$.

Define $d := \frac{1}{2} - e$.

For each positive rational number r we define $U_r := [\frac{1}{2} - r, \frac{1}{2} + r] \times [\frac{1}{2} - r, \frac{1}{2} + r]$. Note that U_r is a square with midpoint $(\frac{1}{2}, \frac{1}{2})$ and with sides of length $2r$, and that $U = U_{\frac{1}{2}}$.

For each positive rational number r we let τ_r be an element of $\mathcal{N}$ enumerating a continuous function from U to U_r such that for all x, y in $[-\frac{1}{2}, \frac{1}{2}]$, $\tau_r(\frac{1}{2} + x, \frac{1}{2} + y) = (\frac{1}{2} + 2rx, \frac{1}{2} + 2ry)$.

For each positive rational number $r < \frac{1}{2}$ we let σ_r be an element of $\mathcal{N}$ enumerating a continuous function from U to U such that, for all x, y in $[-\frac{1}{2}, \frac{1}{2}]$,

(1) if $r \leqslant sup(|x|, |y|) \leqslant \frac{1}{2}$, then $\sigma_r(\frac{1}{2} + x, \frac{1}{2} + y)) = \langle \frac{1}{2} - \frac{x}{2sup(|x|,|y|)}, \frac{1}{2} - \frac{y}{2sup(|x|,|y|)} \rangle$, and,
(2) if $sup(|x|, |y|) \leqslant \frac{r}{2}$, then $\sigma_r(\frac{1}{2} + x, \frac{1}{2} + y) = \langle \frac{1}{2} + x, \frac{1}{2} + y \rangle$, and,
(3) if $\frac{r}{2} \leqslant sup(|x|, |y|) \leqslant r$, then $\sigma_r(x, y) = \langle \frac{1}{2} + \frac{3r - 4sup(|x|,|y|)}{r} x, \frac{1}{2} + \frac{3r - 4sup(|x|,|y|)}{r} y \rangle$.

Note that σ_r maps $U \setminus U_r$ onto ∂U and that σ_r leaves $U_{\frac{r}{2}}$ invariant. Also note that σ_r reflects ∂U through the origin in itself and that, for every point p in $U \setminus U_r$, $|\sigma_r(p) - p| \geqslant \frac{1}{2} + r$.

Now let ϕ be an element of $\mathcal{N}$ enumerating a continuous function from U to U with the property that, for every point p in ∂U, $\phi(p) = p$.

Note τ_d maps U onto U_d and that, under this mapping, interior points of U correspond to interior points of U_d. Also note that $\tau_{\frac{1}{4d}}$ maps U_d onto $U = U_{\frac{1}{2}}$ and that, under this mapping, interior points of U_d correspond to interior points of U. Observe that, for every p in ∂U_d, $\tau_d(\phi(\tau_{\frac{1}{4d}}(p)) = p$.
Finally, note that σ_d maps U onto U and that, under this mapping, interior points of U_d correspond to interior points of U.

We let ψ be an element of $\mathcal{N}$ enumerating a continuous function from U to U such that, for every point p in U, if p belongs to U_d, then $\psi(p) = \sigma_d\big(\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))\big)$, and if p does not belong to U_d, then $\psi(p) = \sigma_d(p)$.

Using (iii), find a point p in U such that $|\psi(p) - p| < e$. Note that p must belong to U_d, as, for every q in $U \setminus U_d$, $|\sigma_d(q) - q| \geqslant \frac{1}{2} > e$, and, therefore, $|\psi(q) - q| > e$.

We may make the following case distinction: *either* $\psi(p)$ is an interior point of U *or* $\psi(p)$ is apart from $(\frac{1}{2}, \frac{1}{2})$.

Let us first assume that $\psi(p)$ is an interior point of U. Note that $\psi(p) = \sigma_d\big(\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))\big)$. We conclude that $\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))$ is an interior point of U_d and that $\phi(\tau_{\frac{1}{4d}}(p))$ is an interior point of U. So we have found a point $q = \tau_{\frac{1}{4d}}(p)$ in U such that $\phi(q)$ is an interior point of U.

Let us now assume that $\psi(p) = \sigma_d\big(\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))\big)$ is apart from the midpoint $(\frac{1}{2}, \frac{1}{2})$ of U. Considering the definition of σ_d, we conclude that also $\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))$ is apart from $(\frac{1}{2}, \frac{1}{2})$. Find q on the border ∂U_d of U_d such that q is on the straight line through $(\frac{1}{2}, \frac{1}{2})$ and $\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))$ and on the same side of $(\frac{1}{2}, \frac{1}{2})$ as $\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))$. Note that $\sigma_d(q)$ belongs to the border ∂U of U and that the straight line connecting the points $\sigma_d(q)$ and p passes through the midpoint $(\frac{1}{2}, \frac{1}{2})$ of U, and that, on this line, the points $\sigma_d(q)$ and p are lying on opposite sides of the point $(\frac{1}{2}, \frac{1}{2})$. Therefore, $|\sigma_d(q) - p| > \frac{1}{2} > e$. As $|\psi(p) - p| < e$, $\sigma_d(q)$ lies apart from $\psi(p) = \sigma_d\big(\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))\big)$. It follows that q lies apart from $\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))$, so $\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))$ is an interior point of U_d and $\phi(\tau_{\frac{1}{4d}}(p))$ is an interior point of U. So again we have found a point $r = \tau_{\frac{1}{4d}}(p)$ in U such that $\phi(r)$ is an interior point of U.

(iv) $\Rightarrow$ (i). Suppose that β belongs to $\mathcal{C}$ and that D_β is bar in $\mathcal{C}$. We have to prove that there exists n such that $D_{\overline{\beta}(n)}$ is a bar in $\mathcal{C}$.

We first define a function B from the set of all binary sequence numbers to the set of the rational segments included in [0, 1], by induction on the length of the argument, as follows:

(i) $B(\langle\ \rangle) = \langle 0, 1\rangle$.
(ii) For every binary sequence number b, for all p, q, if $B(b) = \langle p, q\rangle$, then $B(b * \langle 0\rangle) = \langle p, \frac{1}{2}(p+q)\rangle$ and $B(b * \langle 1\rangle) = \langle \frac{1}{2}(p+q), q\rangle$.

For each n, we let C_n be the set of all rational segments s, such that for some $k < n$, k belongs to $D_{\overline{\beta}(n)}$ and no proper initial part of k belongs to $D_{\overline{\beta}(n)}$ and $s = B(k)$.

Using Lemma 4 we now construct ϕ in $\mathcal{N}$ such that

(1) for each n, ϕ^n enumerates a partial continuous function from $\overline{C_n} \times \overline{C_n}$ to $\mathbb{R}$,
(2) for each n, ϕ^{n+1} extends ϕ^n,
(3) for each n, if $\overline{C_n}$ is a proper subset of [0, 1], then ϕ^n maps $\overline{C_n} \times \overline{C_n}$ to ∂U and for every p in ∂U, if p belongs to $dom_{\mathbb{R}^2}(\phi^n)$, then $\phi^n(p) = p$, and
(4) for all n, if n is the first k such that $\overline{C_k}$ coincides with [0, 1], then ϕ_n is a function from U to U such that, for every p in U, if p belongs to ∂U, then $\phi_n(p) = p$, and, for all j, $\phi_{n+j} = \phi_n$.

Observe that $E_\phi = \bigcup_{n\in\mathbb{N}} E_{\phi^n}$. We conclude that ϕ enumerates a partial continuous function from $\bigcup_{n\in\mathbb{N}}(\overline{C_n} \times \overline{C_n})$ to $\mathbb{R}$.

Let x be an element of [0, 1] such that either x coincides with a dyadic rational or x is apart from every dyadic rational. Find γ in $\mathcal{C}$ such that, for every i, x belongs to $B(\overline{\gamma}i)$. Find n such that $\beta(\overline{\gamma}n) = 1$ and conclude that x belongs to $\overline{C_n}$. It follows that every point (x, y) in U such that both x and y either coincide with a dyadic rational or are apart from every dyadic rational belongs to $dom_{\mathbb{R}^2}(\phi)$. Using Lemma 5 we conclude that $dom_{\mathbb{R}^2}(\phi)$ coincides with U and that ϕ enumerates a continuous function from U to U such that, for every p in ∂U, $\phi(p) = p$. Using (iii), we determine p such that $\phi(p)$ is an interior point of U. We then determine q in U such that both q^0 and q^1 coincide with dyadic rationals and $\phi(q)$ is an interior point of U, and we determine n such that both q^0 and q^1 belong to $\overline{C_n}$. It now follows from (3) that $\overline{C_n}$ is not a proper subset of [0, 1], and, therefore, $D_{\overline{\beta}n}$ is a bar in $\mathcal{C}$. □

Theorem 8 *The following statements are equivalent in* BIM.
(i) Kleene's Alternative to the Weak Fan Theorem: there is a decidable subset of $\mathbb{N}$ *that is a bar in* $\mathcal{C}$*, while every one of its finite subsets positively fails to be a bar in* $\mathcal{C}$*.*
(ii) There exists ϕ *enumerating a continuous function from* U *to* ∂U *such that, for every* p *in* ∂U*,* $\phi(p) = p$*.*
(iii) For all positive rational numbers e *such that* $e < \frac{1}{2}$*, there exists* ϕ*, enumerating a continuous function from* U *to* U *such that, for every* p *in* U*,* $|\phi(p) - p| \geqslant e$*.*
(iv) There exists a positive rational number e *and there exists* ϕ*, enumerating a continuous function from* U *to* U *such that, for every* p *in* U*,* $|\phi(p) - p| \geqslant e$*.*

Proof (i) ⇒ (ii). Let β be an element of $\mathcal{C}$ such that D_β is a bar in $\mathcal{C}$ while, for each n, $D_{\overline{\beta}n}$ is not a bar in $\mathcal{C}$. We now construct ϕ enumerating a continuous function from U to U just as we did in the proof of Theorem 6 (iv)⇒(i). Note that ϕ enumerates a continuous function from U to ∂U such that, for every p in ∂U, $\phi(p) = p$.
(ii) ⇒ (iii). Let ϕ be an element of $\mathcal{N}$ enumerating a continuous function from U to ∂U such that, for every p in ∂U, $\phi(p) = p$.

Now let e be a positive rational number such that $e < \frac{1}{2}$. We define $d := \frac{1}{2} - e$ and, using the functions we defined in the proof of Theorem 7 (iii)⇒(iv), we let ψ be an element of $\mathcal{N}$ enumerating a continuous function from U to U such that, for every point p in U, if p belongs to U_d, then $\psi(p) = \sigma_d\big(\tau_d(\phi(\tau_{\frac{1}{4d}}(p)))\big)$, and if p does not belong to U_d, then $\psi(p) = \sigma_d(p)$.

We claim that ψ satisfies the requirements:

Let p belong to U. By the argument given in the proof of Theorem 7, we know that, if $|\psi(p) - p| < e$, then there exists q in U such that $\phi(q)$ is an interior point of U. As there is no such q, $|\psi(p) - p| \geqslant e$.

(iii) $\Rightarrow$ (iv). Obvious.

(iv) $\Rightarrow$ (i). Let e be a positive rational number and let ϕ enumerate a continuous function from U to ∂U such that, for every p in U, $|\phi(p) - p| \geqslant e$. We show that the function enumerated by ϕ positively fails to be uniformly continuous, and thus obtain Kleene's Alternative, using the observation we made at the end Section 3.

Let n be a natural number such that $\frac{1}{2^n} < \frac{1}{6}e$. As in the proof of Theorem 6 we consider all points of the form $(\frac{p}{2^{n+1}}, \frac{q}{2^{n+1}})$, where $p, q \leqslant 2^{n+1}$ and we label them as we did in the earlier argument. We again find an elementary triangle pqr such that its vertices p, q, r receive the three labels 0, 1, 2. Take any point (x, y) contained in this triangle. Note that we can make six decisions: for each $i < 2$, for each $z = (z_0, z_1)$ in $\{p, q, r\}$, *either* $|\phi^i(x, y) - \phi^i(z_0, z_1))| < \frac{1}{6}e$ *or* $|\phi^i(x, y) - \phi^i(z_0, z_1))| > \frac{1}{12}e$. Suppose we choose six times the first alternative. Reasoning as before, we obtain the conclusion: $|\phi(x, y) - (x, y)| < e$. Contradiction. Therefore, we are sure to choose at least one time the second alternative and thus find points (x, y) and (z_0, z_1) such that $|(x, y) - (z_0, z_1)| < \frac{1}{2^n}$ and $|\phi^i(x, y) - \phi^i(z_0, z_1))| > \frac{1}{12}e$ and, therefore, also $|\phi(x, y) - \phi(z_0, z_1))| > \frac{1}{12}e$.

Thus we see how to find, given any n, points r, t in U such that $|r - t| < \frac{1}{2^n}$ and $|\phi(r) - \phi(t)| > \frac{1}{12}e$. Clearly then, the function enumerated by ϕ positively fails to be uniformly continuous. □

A few final observations are perhaps in order.

In the first place, our results should be compared to the results obtained in the classical Reverse Mathematics program initiated by H. Friedman and S. Simpson. The reader may consult Theorem IV.7.7 in [12]. It follows from this theorem that Weak König's Lemma is equivalent, in RCA_0, to the statement that every continuous function from U to U has a fixed point. These two statements both fail to be true constructively. Theorem 7 establishes that the Weak Fan Theorem, from the classical point of view a contrapositive formulation of Weak König's Lemma, is equivalent, in BIM, to the statement that every continuous function from U to U has approximate fixed points. Our argument may be used for proving that, in RCA_0, the latter statement is also an equivalent of Weak König's Lemma, a fact that we did not find in [12].

Secondly, the reader may be surprised by the equivalences we proved and ask for some deeper explanation. We have not much to say in this respect and refer the reader to [13], where other surprising equivalents both of the Weak Fan theorem and of Kleene's Alternative are brought to light.

Thirdly, J. Berger and H. Ishihara prove, in [1], that the fan theorem is equivalent to the following statement, called $FIX!$:

Each uniformly continuous function Φ from a compact metric space X into itself with at most one fixed point and approximate fixed points has a fixed point.

Here Φ is said to have *at most one fixed point*, if for each p, q in X such that $p \# q$, either $\Phi(p) \# p$ or $\Phi(q) \# q$. Combining this observation with Theorem 6 we may conclude:

The fan theorem implies that each uniformly continuous function Φ from U into itself with at most one fixed point has a fixed point.

This observation made us formulate and prove the following result.

Theorem 9 The following statements are equivalent in BIM.
(i) The Weak Fan Theorem.
(ii) Brouwer's fixed-point theorem with a uniqueness hypothesis: every enumerable continuous function from U into itself with at most one fixed point has a fixed point.

Proof (i) $\Rightarrow$ (ii). One may consult [1] and follow the argument sketched just before this theorem. A direct proof may be obtained as follows. Suppose ϕ enumerates a continuous function from U into itself with at most one fixed point. Using the Weak Fan Theorem, one may prove that, for any two positively disjoint compact subsets V, W of U, there exists a positive rational e such that *either*, for all p in V, $|\phi(p) - p| > e$, *or*, for all p in W, $|\phi(p) - p| > e$. Using this, one may construct a point r in U and two sequences $d_0, d_1, \ldots$ and $e_0, e_1, \ldots$ of positive rationals such that $\lim\limits_{n \to \infty} d_n = 0$ and, for all n, for all q in U, if $|r - q| > d_n$, then $|\phi(q) - q| > e_n$. It now follows that $\phi(r) = r$. For suppose $|\phi(r) - r| > 0$. Using the fact that the function enumerated by ϕ is continuous at r, we find a positive rational number e such that, for all p in U, $|\phi(p) - p| > e$. It follows from The Weak Fan Theorem that the function enumerated by ϕ is uniformly continuous, and, by Theorem 6, we have a contradiction.
(ii) $\Rightarrow$ (i) Let β be an element of $\mathcal{C}$ such that D_β is a bar in $\mathcal{C}$. We want to find a natural number n such that $D_{\overline{\beta}n}$ is a bar in $\mathcal{C}$.

As in the proof of Theorem 7(iv) $\Rightarrow$ (i), we define a function B from the set of all binary sequence numbers to the set of the rational segments included in [0, 1], by induction on the length of the argument, as follows:

(i) $B(\langle\ \rangle) = \langle 0, 1\rangle$.
(ii) For every binary sequence number b, for all p, q, if $B(b) = \langle p, q\rangle$, then $B(b * \langle 0\rangle) = \langle p, \frac{1}{2}(p+q)\rangle$ and $B(b * \langle 1\rangle) = \langle \frac{1}{2}(p+q), q\rangle$.

Again, for each n, we let C_n be the set of all rational segments s, such that for some $k < n$, k belongs to $D_{\overline{\beta}(n)}$ and no proper initial part of k belongs to $D_{\overline{\beta}(n)}$ and $s = B(k)$.

Using Lemma 4 we now construct ψ in $\mathcal{N}$ such that

(1) for each n, ψ^n enumerates a partial continuous function from $\overline{C_n} \times \overline{C_n}$ to $\mathbb{R}$,
(2) for each n, ψ^{n+1} extends ψ^n,

(3) for each n, if $\overline{C_n}$ is a proper subset of $[0, 1]$, then ψ^n maps $\overline{C_n} \times \overline{C_n}$ to ∂U and for every p in ∂U, if p belongs to $dom_{\mathbb{R}^2}(\phi^n)$, then $\psi^n(p) = \psi^n(\langle p^0, p^1\rangle) = -p = \langle -p^0, -p^1\rangle$, and
(4) for all n, if n is the first k such that $\overline{C_k}$ coincides with $[0, 1]$, then ψ_n is a function from U to U *with exactly one fixed point*, that is, there exists r in U such that $\psi^n(r) = r$, and, for all q in U, if $q\#r$, then $\psi^n(q)\#q$, with the additional property that, for every p in U, if p belongs to ∂U, then $\psi_n(p) = -p$. We also require that, for all j, $\psi_{n+j} = \psi_n$.

We leave it to the reader to verify that, for each n, if $\overline{C_n}$ is a proper subset of $[0, 1]$, then, for all p in $dom_{\mathbb{R}^2}(\psi^n)$, $\psi^n(p) \# p$.

We guess that the reader will be able to define ψ^n in case (4). The idea would be to choose the intended fixed point r in $U \setminus \overline{C_{n-1}} \times \overline{C_{n-1}}$ and to obtain ψ^n by linear interpolation from ψ^{n-1} and the (very) partial function with domain $\{r\} \cup \partial U$ mapping r onto r and every point p in ∂U onto $-p$.

Observe that $E_\psi = \bigcup\limits_{n\in\mathbb{N}} E_{\psi^n}$. We conclude that ψ enumerates a partial continuous function from $\bigcup\limits_{n\in\mathbb{N}} (\overline{C_n} \times \overline{C_n})$ to $\mathbb{R}$. As in the analogous case in the proof of Theorem 7(iv) $\Rightarrow$ (i), we may conclude that ψ maps U into U.

Moreover, ψ has at most one fixed point. For suppose p, q belong to U and $p \# q$. Find n such that both p and q belong to $\overline{C_n} \times \overline{C_n}$. Note that, if $\overline{C_n}$ is a proper subset of $[0, 1]$, then both $\psi(p) \# p$ and $\psi(q) \# q$. If, on the other hand, $\overline{C_n}$ coincides with $[0, 1]$, then $E_\psi = E_{\psi^n}$ and we may find r such that $\psi^n(r) = r$ and, for all z, if $z \# r$, then $\psi^n(z) \# z$. As either $p \# r$ or $q \# r$, we are done.

Using (iii), we determine z in U such that $\psi(z) = z$. Find n such that z belongs to $\overline{C_n} \times \overline{C_n}$. Note that $\psi(z)$ coincides with $\psi^n(z)$. Recall that, if $\overline{C_n}$ is a proper subset of $[0, 1]$, then, for all p in $dom_{\mathbb{R}^2}(\psi^n)$, $\psi^n(p) \# p$. It follows that $\overline{C_n}$ coincides with $[0, 1]$ and that $D_{\overline{\beta}n}$ is a bar in $\mathcal{C}$. □

Like Theorem 7, this theorem has its counterpart.

Theorem 10 The following statements are equivalent in BIM.
(i) Kleene's Alternative to the Weak Fan Theorem.
(ii) There exists ϕ enumerating a continuous function from U to itself such that, for each p in U, $\phi(p) \# p$.

Proof (i) $\Rightarrow$ (ii). See Theorem 8(i) $\Rightarrow$ (iii).
(ii) $\Rightarrow$ (i). Let ϕ be an element of of $\mathcal{N}$ enumerating a continuous function from U to itself such that, for each p in U, $\phi(p) \# p$. We now let ψ enumerate a continuous function from U to itself such that, for each p that is an interior point of U, $\psi(p)$ is the intersection point of ∂U and the line connecting p and $\phi(p)$ that lies on the same side of $\phi(p)$ as p, and, for each p in ∂U, $\psi(p) = p$. Note that ψ maps U onto ∂U and that, for each p in ∂U, $\psi(p) = p$. Kleene's Alternative follows by Theorem 8(ii) $\Rightarrow$ (i). □

Note that the statement (ii) in Theorem 10 may be construed as a constructive denial of statement (ii) in Theorem 9: *there exists ϕ enumerating a continuous func-*

tion from U to itself that has at most one fixed point and positively fails to have a fixed point.

The argument used for Theorem 10(ii) $\Rightarrow$ (i) is sometimes mentioned as an intuitively clear way to prove (classically, by contraposition) the two-dimensional case of Brouwer's fixed point theorem from the intuitively obvious (?) fact Theorem 7(iv), see [8], Section V.4.

Acknowledgement We want to thank the referee for his careful reading of the paper and for making some useful observations.

References

1. J. Berger and H. Ishihara, Brouwer's fan theorem and unique existence in constructive analysis, *Math. Logic Quart.* 51(2005), pp. 360–364.
2. L.E.J. Brouwer, Über Abbildungen von Mannigfaltigkeiten, Mathematische Annalen 71(1911), pp. 97–115, also in [7], pp. 454–472.
3. L.E.J. Brouwer, Beweis dass jede volle Funktion gleichmässig stetig ist, *Koninklijke Nederlandse Akademie van Wetenschappen Verslagen* 27(1924), pp. 189–193, also in: [6], pp. 286–290.
4. L.E.J. Brouwer, An intuitionistic correction of the fixed-point theorem on the sphere, *Proc. Roy. Soc. London* , Ser. A 213(1952), pp. 1–2, also in [6], pp. 506–507.
5. L.E.J. Brouwer, Door klassieke theorema's gesignaleerde pinkernen die onvindbaar zijn, *Indag. math.* 14(1952), pp. 443–445, translation: Fixed cores which cannot be found, though they are claimed to exist by classical theorems, in [6], pp. 519–521.
6. L.E.J. Brouwer, *Collected Works, Vol. I: Philosophy and Foundations of Mathematics*, ed. A. Heyting, North Holland Publ. Co., Amsterdam, 1975.
7. L.E.J. Brouwer, *Collected Works, Vol. II: Geometry, Analysis, Topology and Mechanics*, ed. H. Freudenthal, North Holland Publ. Co., Amsterdam, 1976.
8. R. Courant, H. Robbins, *What is Mathematics? An Elementary Approach to Ideas and Methods*, Oxford University Press, Oxford, 1941, new edition, revised by I. Stewart, 1996.
9. S.C. Kleene, R.E. Vesley, *The Foundations of Intuitionistic Mathematics, Especially in Relation to the Theory of Recursive Functions*, Amsterdam, North-Holland Publ. Co., 1965.
10. V.P. Orevkov, A constructive mapping from the square onto itself displacing every constructive point, *Soviet Math. Doklady* 4(1963), pp. 1253–1256.
11. N. Shioji, K. Tanaka, Fixed point theory in weak second-order arithmetic, *Ann. Pure App. Logic*, 47(1990), pp. 167–188.
12. S.G. Simpson, *Subsystems of Second Order Arithmetic*, Perspectives in Mathematical Logic, Springer Verlag, Berlin etc., 1999.
13. W. Veldman, *Brouwer's Fan Theorem as an axiom and as a contrast to Kleene's alternative*, Report No. 0509, Department of Mathematics, Faculty of Science, Radboud University Nijmegen, July 2005.
14. W. Veldman, Brouwer's Real Thesis on Bars, *Philosophia Scientiae*, Cahier Spécial 6(2006) 21–39.
15. W. Veldman, *Some consequences of Brouwer's thesis on bars*, in: One Hundred years of Intuitionism (1907–2007), The Cerisy Conference, ed. M. van Atten, P. Boldini, M. Bourdeau, G. Heinzmann, Birkhäuser, Basel etc., 2008, pp. 326–340.

Part III
Formalism

“Gödel’s Modernism: On Set-Theoretic Incompleteness,” Revisited

Mark van Atten and Juliette Kennedy

As to problems with the answer Yes or No, the conviction that they are always decidable remains untouched by these results.

—Gödel

1 Introduction

1.1 Questions of Incompleteness

On Friday, November 15, 1940, Kurt Gödel gave a talk on set theory at Brown University.[1] The topic was his recent proof of the consistency of Cantor’s Continuum Hypothesis, henceforth *CH*,[2] with the axiomatic system for set theory

M. van Atten (✉)
Centre National de Recherche Scientifique, Paris
e-mail: Mark.van Atten@hniv-paris1.fr

An earlier version of this paper appeared as ‘Gödel’s modernism: on set-theoretic incompleteness’, *Graduate Faculty Philosophy Journal*, 25(2), 2004, pp. 289–349. Erratum facing page of contents in 26(1), 2005.

[1] All of Gödel’s published papers and selections from his unpublished papers and correspondence have been published in five volumes: K. Gödel, *Collected Works*, eds. S. Feferman et al. (Oxford: Oxford University Press). *I: Publications 1929–1936* (1986); *II: Publications 1938–1974* (1990); *III: Unpublished essays and lectures* (1995); *IV: Correspondence A-G* (2003); *V: Correspondence H-Z* (2003). We will refer to these volumes as *CW I*, etc. The Brown University lecture is in *CW III:175–185*.

[2] *CH* is the hypothesis that $2^{\aleph_0} = \aleph_1$. *GCH* is its generalization $2^{\aleph_\alpha} = \aleph_{\alpha+1}$. Georg Cantor had first stated a weaker form of *CH* in 1878 and then *CH* in 1883. *GCH* was formulated by Felix Hausdorff in 1908. For the history of *(G)CH*, see in particular J. Dauben, *Georg Cantor. His Mathematics and Philosophy of the Infinite* (Princeton: Princeton University Press, 1990), M. Hallett, *Cantorian Set Theory and Limitation of Size* (Oxford: Clarendon Press, 1984), and the introductions to the relevant papers in K. Gödel, *Collected Works. II: Publications 1938–1974*, eds. S. Feferman et al. (Oxford: Oxford University Press, 1990) and K. Gödel, *Collected Works. III: Unpublished essays and lectures*, eds. S. Feferman et al. (Oxford: Oxford University Press, 1995). We will henceforth refer to these volumes by their title and number only, e.g., *Collected Works II*. We will sometimes refer to papers by their year, following the system of *CW*; the year of unpublished papers is preceded by an asterisk, there is a letter suffix in case different papers appeared in the same year (e.g., **1940a*), and there is a question mark if the year is not certain.

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_15, © Springer Science+Business Media B.V. 2009

ZFC.[3] His friend from their days in Vienna, Rudolf Carnap, was in the audience, and afterward wrote a note to himself in which he raised a number of questions on incompleteness:[4]

> (Remarks I planned to make, but *did not*)
>
> Discussion on Gödel's lecture on the Continuum Hypothesis, November 14,[5] 1940
>
> There seems to be a difference: between the *undecidable* propositions of the kind of his example [i.e., 1931] and propositions such as *the Axiom of Choice*, and *the Axiom of the Continuum* [*CH*].
>
> We used to ask: "When these two have been decided, is then everything decided?" (The Poles, Tarski I think, suspected that this would be the case.) Now we know that (on the basis of *the usual finitary rules*) there will always remain undecided propositions.
>
> 1. Can we nevertheless still ask an analogous question? i.e. is there an objective difference between 2 kinds of problems, or is it just a difference in degree of simplicity?
> 2. If so, are there grounds for a positive answer? i.e., "Now that we have accepted both axioms, all simple problems are determined?"

We recapitulate the basic facts. In 1931, Gödel proved his well-known theorem: for every ω-consistent formal system that contains arithmetic and is recursively axiomatizable, as we would say now, there exist sentences ϕ (in the language of the system) such that neither ϕ nor $\neg\phi$ is derivable in the system. Such a sentence is said to be undecidable in the system and renders it incomplete. The three conditions on a formal system mentioned in the theorem mean the following. 1. ω-consistency means that the system should not prove (for some P definable in it) $\exists x \neg P(x)$ while also proving $P(n)$ for each natural number term n. 2. Containing a sufficient amount of arithmetic means that the operations of addition, multiplication, successor, as well as the notion of an order, should be definable in the system, and that the principle of induction should be included. 3. Recursive axiomatizability means that the axioms should be either finite in number or enumerable by an effective procedure. (In 1936, J.B. Rosser showed that the requirement of ω-consistency can be weakened to consistency.)

The class of formal systems to which the incompleteness theorem applies includes all of the more ambitious formal systems that had been formulated up till 1931: Principia Mathematica, the systems devised by Hilbert and his followers, and, in particular, the system of set theory that is still the canonical system today, ZFC.

For more general histories of set theory, see A. Kanamori, "The Mathematical Development of Set Theory from Cantor to Cohen," *The Bulletin of Symbolic Logic* 2(1) (1996), pp.1–71, and J. Ferreiros Dominguez, *Labyrinth of Thought: A History of Set Theory and Its Role in Modern Mathematics* (Boston: Birkhäuser, 1999).

[3] ZFC is Zermelo-Fraenkel set theory with the Axiom of Choice.

[4] *Kurt Gödel. Wahrheit und Beweisbarkeit. Band 1: Dokumente und historische Analysen.*, eds. E. Köhler et al. (Wien: öbv & hpt, 2002), pp. 127–128; translation from the German ours.

[5] According to *CW I*, p. 41, the Brown lecture was on November 15, not 14.

Although the theorem shows that, for each system of the type described, there are undecidable sentences, it does not show that there is a sentence that cannot be decided in any possible system of that type. However, the theorem does not exclude the existence of such a sentence either. If it exists, it could be called absolutely undecidable (we will introduce a slightly more refined terminology below).[6]

In this paper, we will be concerned with incompleteness and undecidability in ZFC and related systems for set theory. The question that will be in the foreground most, and in the background all of the time, is: do absolutely undecidable propositions exist in set theory? We will analyse specifically how Gödel's thinking about this question developed in his published and unpublished work, closing with considerations on the present situation in set theory in the light of Gödel's ideas.

1.2 Splitting of the Notion of Undecidability

Gödel held that the axioms of classical Zermelo-Fraenkel set theory (or some system equivalent to it) are true and evident. It must have been they that he had in mind when he said, in 1966, that the axiomatization of set theory was the greatest advance in its foundations prior to forcing.[7] But they cannot be more than an initial segment of the correct axioms for all of mathematics, as by the incompleteness theorems, there are sentences ϕ (in the language of ZFC) that are undecidable in ZFC. With Carnap (see above), one can ask whether the collection of undecidable sentences is exhausted by those constructed in the proof of the incompleteness theorem, and this question is central to the present paper.

Clearly, any ϕ undecidable in ZFC falls into at least one of the following nominally defined categories, which split the notion of undecidability:

1. sentences that are undecidable in ZFC but seen to be true (and hence decided informally) by reflecting on the proof of their undecidability in ZFC.
2. sentences that are undecidable in ZFC, and are not decided informally by reflecting on the proof.
3. sentences that are undecidable in ZFC, but are decidable in an evident extension (or series of extensions) of ZFC.

---

[6] By their characteristic logic based on their characteristic notion of truth, according to intuitionists there are no absolutely undecidable propositions. For assume that ϕ is absolutely undecidable. Then in particular the assumption that ϕ has been proved must lead to a contradiction. But if it does, this, on the intuitionistic conception of negation, is to say that $\neg\phi$ holds, which decides ϕ and thereby contradicts the assumption. But by the same idiosyncracies of intuitionistic logic, this little argument does not show that therefore every proposition is decidable. In fact intuitionists consider that as highly unlikely, given that on their interpretation that would mean that one knows a universal method to decide all mathematical propositions.

[7] *CW II*, pp. 269-70

4. sentences that are undecidable in ZFC, are not decidable in any evident extension of ZFC, but can be decided by human reason.
5. sentences that are undecidable in ZFC, are not decidable in any evident extension of ZFC, and cannot be decided by human reason.

These categories are not all mutually exclusive, for example ϕ may be of both the first and third category, or of the second and the third (if one finds a new axiom by other means than reflecting on undecidability proofs), or of the second and fifth. The questions at hand are the following: of which of these five categories, if any, can we establish that they are not empty? And if a category is not empty, do its members admit of a systematic characterization? It is crucial here that "extension" is taken in a non-trivializing sense: one adds only axioms that are seen to be true or evident. Simply adding ϕ without considering its evidence would miss the point. Note that an extension of a formal system may also consist in, or also involve, adding higher types to the logic or otherwise changing the logic in some appropriate way.[8]

Category 4 seems to be necessarily empty. For, on any reasonable informal understanding of proof, a proof of a sentence (or of its negation) proceeds from evident axioms, by evident inferences, to its conclusion; it is, as Gödel put it, 'not … a sequence of expressions satisfying certain formal conditions, but a sequence of thoughts convincing a sound mind'.[9] So conversely, for any mathematical sentence that human reason decides, it should be able to indicate the evident axioms and evident inferences on the basis of which the decision was made. But if this can be done (i.e., if oracles are not admitted), these can be formalized and used to extend ZFC. (Note that 'arguments from success' only lead to probable decisions, and such arguments are therefore not excluded by the emptiness of category 4.)

To demonstrate the non-emptiness of category 1, we can simply use an undecidable statement constructed along the lines of Gödel's proof of the incompleteness theorem. By theorems of Gödel (1938) and Cohen (1963) that will play an important role in this paper, one can take $\phi = CH$ to give an example of a statement in category 2 (and hence not in category 1). But it is at present not known whether CH exemplifies the non-emptiness of category 3 or 5 (excluding 4 for the reason given above). It must be in one of them, and, as 3 and 5 are disjoint, exactly one. At the end of this paper we will consider the suggestion associated with the Woodin school to the effect that CH could be solved. What can be said about category 5 will depend on how strong and specific one's views are on the nature of reason as well as on the ontology of mathematics.[10]

---

[8] In an explicitly typed system, adding higher types means that, besides variables for individuals (e.g. numbers, or sets), there will also be variables for sets of individuals, sets of sets of individuals, and so on, together with appropriate axioms that govern formation of sets of these types. In the cumulative hierarchy of sets, adding higher types is just adding more levels to the hierarchy.

[9] *CW III*, p. 341n.20.

[10] See the remarkable reflections on the topic by Emil Post in the appendix to a manuscript from 1941 titled "Absolutely Unsolvable Problems and Relatively Undecidable Propositions," first published in *The Undecidable*, ed. M. Davis (Hewlett NY: Raven Press, 1965), pp. 340–433; the appendix starts on p. 41.

A distinction that cuts across this classification of statements undecidable in ZFC into five categories is that between statements that do play a role in mathematical practice and those that do not. This may of course change through time and therefore unlike the five-fold classification this one is not fixed. To consider these two distinctions in tandem is motivated by the fact that the undecidable sentences constructed in the proofs of Gödel's incompleteness theorems are manifestly different from anything found in mathematical practice so far, and in that specific sense not mathematically meaningful; we will take this specific sense as our definition of mathematical meaningfulness. The greatest interest is in the question whether a statement can be found that is both mathematically meaningful and absolutely undecidable, for that would make urgent the search for a new evident axiom from a practical perspective.

To a realist, the mathematical meaningfulness of a statement simply means that it has mathematical content (or is equivalent to one that does), in the sense that the terms in the statement refer.[11] Such statements can be called "contentual" ("inhaltlich"). To a (Hilbertian) formalist a certain statement may well be relevant to mathematics without being contentual (think of any practically relevant part of classical mathematics that is not finitary). On the other hand, this kind of background commitment associated with the realist and the formalist is often lacking in the the colloquial use of the phrase "mathematically meaningful" among mathematicians. That use rather emphasizes typical aspects (often of an aesthetic nature) such as being "natural," "fundamental," "elementary," or "interesting."

The question as to the cardinality of the continuum, a decision of *CH*, exemplifies many aspects of mathematical meaningfulness. As Gödel describes it in the 1947 paper, it is one of "the most fundamental questions in the field"; a question "from the "multiplication table" of cardinal numbers."[12]

> So the analysis of the phrase "how many" unambiguously leads to a definite meaning for the question stated in the second line of this paper: The problem is to find out which one of the $\aleph$'s is the number of points of a straight line or (which is the same) of any other continuum (of any number of dimensions) in a Euclidean space.[13]

1.3 *Gödel's View on Undecidability in 1931*

How were incompleteness phenomena understood by Gödel in 1931? Did he expect all undecidable statements to be in category 1, (coded) metamathematical statements (e.g., involving provability, its particular case consistency, or computability) or in

[11] Through the device of contextual definitions, statements may arise that are not meant to be taken at face value, yet are equivalent to ones that should. Gödel **1940a*, *CW III*, p. 176 simply defines meaningfulness as being part of mathematics proper (or translatable into mathematics proper). The relation to practice that we will be concerned with is not thematized there.

[12] *CW II*, p. 257 and p. 256.

[13] *CW II*, p. 257 and p. 256.

any case equivalents of those? Or did he think there are also mathematically meaningful statements which would be in category 2?

In the paper in which Gödel published his incompleteness theorem, he does not go into questions of this type, but in a lecture text which probably is from shortly after (**1931?*), Gödel mentions a concept of "absolute undecidablity" in relation to his theorem:

> The procedure just sketched furnishes, for every system that satisfies the aforementioned assumptions, an arithmetical sentence that is undecidable in that system. That sentence is, however, not at all absolutely undecidable; rather, one can always pass to "higher" systems in which the sentence in question is decidable. (Some other sentences, of course, nevertheless remain undecidable.)[14]

This quotation motivates us to make the following terminological point. As we will see below, at different times Gödel used the term "absolutely undecidable" in different ways. Around 1940 he used it in connection with category 2, but from 1951 onward it refers strictly to category 5 (which is a sub-category of 2). The latter may be the more natural thing to do in any case, for if all we know is that a sentence is of category 2, it is not excluded that we will come to find and believe an axiom that shows the sentence is also of category 3, and the sentence will have been decided after all. No such hope can be entertained if it is somehow shown that a sentence is of category 5, and that circumstance would earn it the predicate "absolutely undecidable" with more justification. We will use "weakly absolutely undecidable" for category 2 and "strongly absolutely undecidable" for category 5.

In the quotation just given, the correct reading of "absolutely undecidable" seems to be "weakly absolutely undecidable," as the reason that Gödel goes on to present is one that distinguishes category 2 from category 1 but does not contain any element that at the same time distinguishes category 5 from its supercategory 2. The reason that Gödel gives is that the higher system in turn is incomplete, and therefore still leaves formally undecided other sentences, which then must have been undecidable in the first system as well. These are decided again in even higher systems, and the story repeats itself, ad infinitum; but it never leads out of category 1.[15] So it seems that Gödel around 1931 mentions (in effect) the notion of weak absolute undecidability only once and in passing.

2 V=L

2.1 1935–1940: A Candidate for Weak Absolute Undecidability

But various people had already begun to entertain the possibility that *CH* may be weakly absolutely undecidable, i.e., that it was not decidable in the known systems for set theory. As early as 1922, Skolem (in a lecture in Helsinki) had conjectured

---

[14] *CW III*, p. 35.
[15] See *CW I*, p. 180, footnote 48a.

that *CH* cannot be decided from the axioms given in Zermelo 1908.[16] Hilbert's well-known attempt in 1925 (published in 1926[17]) to demonstrate *CH* was, as Gregory Moore put it, "met with widespread skepticism, in particular from Fraenkel (1928) and Luzin (1929),"[18] and in Bologna in 1928, Bernays and Tarski discussed with each other the possibility of independence of *CH* from ZFC.[19] The next year, Tarski mentioned this possibility in print; in the closing paragraph of "Geschichtliche Entwicklung und gegenwärtiger Zustand der Gleichmächtigkeitstheorie und der Kardinalzahlarithmetik",[20] he says that, although he does not have any argument to offer, he strongly suspects that *CH* will in the future be shown to be independent from ZF and ZFC. And although Gödel in the lecture **1931?* does not speculate on *CH* being formally undecidable in ZFC,[21] he too may have had it in mind then—to Wang he said in 1976 that "it must have been in the summer of 1930 when [I] began to think about the continuum problem and also heard of Hilbert's proposed solution."[22] But certainly no one at the time was in a position to turn the suspicion of independence into a convincing (partial) argument. This may explain why Gödel mentions the no-

---

[16] T. Skolem, "Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre", in *Selected Works in Logic*, ed. J.E. Fenstad (Oslo: Universitetsforlaget, 1970), p. 149n2.

[17] D. Hilbert, "Über das Unendliche," *Mathematische Annalen* 95 (1926), pp. 161–190.

[18] *CW II*, p. 157. The references are to A. Fraenkel, *Einleitung in die Mengenlehre*, 3rd, revised edition (Berlin: Springer, 1923) and Luzin's talk at the Bologna conference, "Sur les voies de la théorie des ensembles", *Atti del Congresso Internazionale dei Matematici, Bologna 3–10 settembre 1928* (Bologna:Zanichelli), I, pp. 295–299.

[19] We thank Paolo Mancosu for this last detail.

[20] In A. Tarski, *Collected Papers. Vol.1: 1921–1934*, eds. S.R. Givant and R.N. McKenzie (Boston: Birkhäuser, 1986), pp. 233–241. We thank Göran Sundholm for bringing this passage to our attention.

[21] Incidentally, Gödel in 1931 does refer to Hilbert's paper from 1926, but in a different context. See footnote 48a on p. 180/181 of *CW I*.

[22] H. Wang, *Popular Lectures on Mathematical Logic* (New York: Dover, 1993), p. 128. "Jetzt, Mengenlehre," Gödel is alleged to have said around that time ("And now, [on to] set theory"): J. Dawson, *Logical Dilemmas. The Life and Work of Kurt Gödel* (Wellesley: A K Peters, 1997), p. 108n18. Gödel's interest in set theory may have begun to develop as early as 1928 when he requested at the library the volume containing Skolem's talk in Helsinki (mentioned at the beginning of this section), as Dawson notes (p. 120). It is worth adding to this that very likely Gödel did not actually get the volume: the library slip in question has a large question mark at the title of the book, the stamp "Ausleihe" is missing, and the smaller part has not, as would have been usual in case of loan, been detached. This library slip therefore does not constitute evidence that Gödel had seen that book at that time (For a more comprehensive discussion of the archive material and how this corroborates Gödel's statements about his completeness theorem and Skolem, see M. van Atten, "On Gödel's awareness of Skolem's Helsinki lecture" *History and Philosophy of Logic*, 26(4) (2005) pp. 321–326.). Be that as it may, Dawson goes on to mention that in 1930 Gödel requested from libraries various works on (or related to) set theory: Hilbert's list of open problems from 1900 (published in the *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen* of the same year)—of which the first is *CH*—, Fraenkel's *Einleitung in die Mengenlehre* from 1919—in which Gödel will have noticed Fraenkel's skepticism about Hilbert's attempted proof of *CH*—, the text of Skolem's Helsinki lecture again, and von Neumann's papers "Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeine Mengenlehre" and "Die Axiomatisierung der Mengenlehre", both from 1928.

tion of (weakly) absolute undecidability but does not give a (possible) example. He did though have a sense where to look for a partial result, i.e. showing not (as Hilbert had attempted to do) *CH* itself, but its consistency with the axioms of ZFC. He must have arrived at a good idea quickly, for, as Kreisel reports on his conversations with Gödel, "he had the general idea for his proof of *GCH* for *L* as a student."[23]

What Kreisel is referring to is the hierarchy of sets *L* that Gödel was to define in 1935 and which enabled him to establish that, if ZFC is consistent, so is ZFC + *CH* (and, what is more, ZFC+*GCH*). The strategy is the following. One formulation of *CH* is: there are $\aleph_1$ subsets of $\aleph_0$. So one could try to find a restricted notion of set that on the one hand satisfies the axioms of ZFC but on the other is so strict that it allows one to keep count of the subsets generated from every set. This strictness can be given form in a hierarchy that starts, naturally, with the empty set at the bottom level, on top of which, in a controlled way so as not to lose count, higher and higher levels of sets are built out of the ones previously obtained. This hierarchy does perhaps not capture the full notion of set because the notion of set used may be too restricted; but if it is shown that within a model for ZFC one can build the hierarchy *L* and that in this "inner model" *CH* is true, then it has been shown that if ZFC is consistent (i.e., has a model), so is ZFC+*CH*. (Appropriately, Kreisel in his memoir of Gödel gave his section on constructible sets the subtitle "reculer pour mieux sauter".[24]) The consistency proof is relative to ZF; the consistency of ZF itself has not yet been established in the strongest sense of the word.

To obtain a precise definition of such a hierarchy, two fundamental choices have to be made: what ordinals will there be to serve as indices of the subsequent levels in the buildup of the hierarchy, and what is the method to build a higher level from the ones beneath it? For the first, Gödel introduced a notion of predicative definability in first-order logic; impredicative definitions[25] would not respect the idea of constructing the universe from the ground up and thereby make it impossible to count.[26] As for the second question, Gödel told Wang that he experimented "with more and more complex constructions [for obtaining the ordinals needed to built set-theoretical hierarchies] for some extended period between 1930 and 1935."[27] The breakthrough came in 1935 and consisted in the decision simply to take the classical ordinals as given. In particular, this means that one takes the non-definable and non-denumerable ordinals as given. This is a characteristically realist idea and

[23] p. 610 of G. Kreisel, "Review of K. Gödel's Collected works, Vol. II," *Notre Dame Journal of Formal Logic* 31(4) (1990) pp. 602–641.

[24] p. 195 of G. Kreisel, "Kurt Gödel. 28 April 1906–14 January 1978," *Biographical Memoirs of Fellows of the Royal Society* (1980), pp. 149–224.

[25] For a discussion of Gödel's ideas on impredicativity (and of a number of other topics), see W. Tait, "Gödel's Unpublished Papers on Foundations of Mathematics," *Philosophia Mathematica* 9 (2001), pp. 87–126.

[26] For finite sets it makes of course no difference whether one uses this notion of definability or the classical power set as the operation to generate higher levels; one can always simply define a finite set by enumerating its elements. For the infinite case this may well be different (and is in general believed indeed to be.)

[27] Wang, *Popular Lectures*, p. 129.

was what distinguished L from Gödel's earlier efforts at constructing hierarchies of sets.[28] To take the ordinals as given does not detract from the value of the proof, as Gödel explained (in the Brown lecture):

> If you want to use [the set theory based on L] for giving an unobjectionable foundation to mathematics our procedure would of course be preposterous, but for proving the consistency of the continuum hypothesis it is perfectly all right, since what we want to prove is of course only a relative consistency of the continuum hypothesis; i.e., we want to prove its consistency under the hypothesis that set theory, including all its transfinite methods, is consistent. Therefore we are justified in using the whole set theory in the consistency proof (because if a contradiction were obtaind from the continuum hypothesis and if, on the other hand, we could prove its consistency by means of set-theoretical arguments, then these set-theoretical arguments would be contradictory).[29]

We will now be somewhat more precise. Given a set x, a first-order formula $\phi(y, a_1, a_2, \ldots a_n)$ (where all quantifiers range over x) defines a subset of x, namely $\{y \in x | \phi(y, a_1, a_2, \ldots a_n)\}$, where the a_i are specific elements of x that form a (possibly empty) list of parameters. Let $D(x)$ denote the set of all sets thus definable from the set x. Then L is defined as follows (α ranging over all the classically admissible ordinals):

$$\begin{aligned} L_0 &= \emptyset \\ L_\alpha &= D(L_\beta) && \text{if } \alpha = \beta + 1 \\ L_\alpha &= \bigcup_{\beta<\alpha} L_\beta && \text{if } \alpha \text{ is a limit ordinal} \\ L &= \bigcup L_\alpha \end{aligned}$$

The sets that occur at some L_α Gödel called "constructible." (To avoid unintended bewitchment by the terminology, one should keep in mind that this notion goes far beyond what a constructive mathematican would accept.) The idea that every set is constructible, in other words the idea that the universe of all sets V coincides with the collection L, found its formulation in the axiom $V{=}L$.

The use of "V" to refer to the universe of all sets has its origin, via Whitehead and Russell's Principia Mathematica, in Peano; Kreisel reports that what Gödel had meant by "L" was "lawlike."[30] But as Kreisel goes on to say that "at the time [i.e. of the consistency proof] he toyed with the idea that L contained all legitimate definitions of sets," one may also suggest that originally "L" rather stood for the German "legitim (definiert, definierbar)," legitimate in the sense that the definitions

---

[28] For this paragraph, see Wang, *Popular Lectures*, p. 129, and also Gödel's remarks in H. Wang, *A Logical Journey. From Gödel to Philosophy* (Cambridge: MIT Press, 1996), p. 251: 8.1.7, 8.1.8, 8.1.9.

[29] *CW III*, p. 178.

[30] p. 158 of G. Kreisel, "Gödel's Excursions into Intuitionistic Logic," in *Gödel Remembered. Salzburg 10-12 July 1983*, eds. P. Weingartner and L. Schmetterer (Napoli: Bibliopolis, 1987), pp. 65–179

are predicative and in terms of first-order logic. That "lawlike" starts with the same letter would then be a fortunate coincidence of the linguistic kind. (In German, 'law-like' is 'gesetzmäig'.)

Gödel's motivation to look for hierarchies of sets, which eventually led him to L, had been to work on *CH*. But the first result he actually showed about L, after having verified that the ZF axioms hold for it, was that the axiom of choice (*AC*) holds in it: if ZF is consistent, so is ZF+*AC*. He kept this secret at first though he did tell von Neumann when visiting Princeton that year. Also in 1935, Gödel conjectured that $V{=}L \rightarrow CH$ and that therefore *CH* is consistent with ZF and with ZFC. He set out to prove this,[31] but for a long period he struggled with depression and poor health. The proof that $V{=}L \rightarrow (G)CH$ he essentially found during the night of 14 to 15 Juni 1937.[32] On December 15, 1937 he wrote to Karl Menger that he now was trying to prove the independence of *CH* from ZFC (for which, given his earlier result, it would suffice to show that ZFC+¬*CH* is also consistent), but without success so far.[33] He announced his consistency results in print in 1938.[34] He did not mention his expectation of independence, which however he did do in his lecture in Göttingen in 1939.[35] Consistency of ZFC+¬*CH* (and of ZF+¬*AC*) would in fact be established by Paul Cohen in 1963 by a method called forcing.[36] A result by Shepherdson from 1953 made it clear that it is actually impossible to use the method of inner models for ¬*CH* (or ¬*AC*).[37]

In 1938, Gödel claims that

> the consistency proof for A [$V{=}L$] does not break down if stronger axioms of infinity (e.g., the existence of inaccessible numbers) are adjoined to T [or to ZF]. Hence the consistency of A seems to be absolute in some sense, although it is not possible in the present state of affairs to give a precise meaning to this phrase.[38]

This turned out to be only partially correct, as many of the stronger large cardinal axioms that have later been proposed and are believed to be consistent (e.g., measurable cardinals) have been shown to imply $V{\neq}L$; but it is correct for inaccessible, Mahlo and the very large weakly compact cardinals. The reservation Gödel expresses refers to the fact that what was still missing is a way to make exact the

---

[31] K. Menger, "Memories of Kurt Gödel," in his *Reminiscences of the Vienna Circle and the Mathematical Colloquium*, eds. L. Golland, B. McGuinness and A. Sklar (Dordrecht: Kluwer, 1994), p. 214.

[32] He wrote in his notebook "Kont.Hyp. im wesentlichen gefunden in der Nacht zum 14 und 15 Juni 1937' (essentially found the [consistency proof of the generalized] Continuum Hypothesis during the night of 14 to 15 Juni 1937) *CW I*, p. 36, note s.

[33] *CW IV*, pp. 112–115.

[34] *CW II*, pp. 26–27.

[35] *CW III*, p. 155.

[36] P.J. Cohen, "The Independence of the Continuum Hypothesis. I," *Proceedings of the National Academy of Sciences, U.S.A.* 50, pp. 1143–1148.

[37] J.C. Shepherdson, "Inner Models for Set Theory III," *The Journal of Symbolic Logic* 18(2), pp. 145–167. The existence of a minimal model was rediscovered by Cohen in 1963 *CW IV*, p. 376.

[38] *CW II*, pp. 26–27.

notion of the whole transfinite series of possible extensions by axioms of infinity.[39] Could Gödel's reason for the suggestion he made in 1938 have something to do with a passage he in 1972 urged Wang to include in the latter's book *From Mathematics to Philosophy*?

> There used to be a confused belief that axioms of infinity cannot refute the constructibility hypothesis (and therefore even less the continuum hypothesis) since L contains by definition all ordinals. For example, if there are measurable cardinals, they must be in L. However, in L they do not satisfy the condition of being measurable. This is no defect of these cardinals, unless one were of the opinion that L is the true universe. As is well known, all kinds of strange phenomena appear in nonstandard models.[40]

Or had it simply been difficult to imagine the very possibility that large cardinals could be of such a different kind that they violate $V = L$? Indeed, when Dana Scott showed in 1961 that a measurable cardinal (introduced by Ulam in 1930) would do just that, Gödel commented that that is an axiom of infinity "of an entirely new kind," as had become clear only shortly before.[41] In a (draft) letter to Tarski of August 1961, he writes: "You probably have heard of Scott's beautiful result that $V \neq L$ follows from the existence of any such measure for any set. I have not checked this proof either but the result does *not* surprise me."[42] Presumably, this would have surprised him in 1938.

In 1939, Gödel explained his consistency proofs of *AC* and *CH* in a lecture in Göttingen;[43] on that occasion he voiced his suspicion that $V = L$ is strongly absolutely undecidable:

> The consistency of the proposition A (that every set is constructible [$V = L$]) is also of interest in its own right, especially because it is very plausible that with A one is dealing with an absolutely undecidable proposition, on which set theory bifurcates into two different systems, similar to Euclidean and non-Euclidean geometry.[44]

Gödel's remark in Göttingen about Euclidean and non-Euclidean geometry is reminiscent of his remarks in the second edition of the Cantor paper from 1964. There however he makes a comment to quite the opposite effect:

> [I]t has been suggested that, in case Cantor's continuum problem should turn out to be undecidable from the accepted axioms of set theory, the question of its truth would lose its meaning, exactly as the question of the truth of Euclid's fifth postulate by the proof of the consistency of non-Euclidean geometry became meaningless for the mathematician.

---

[39] See also 1946 in *CW II*, p. 151.

[40] H. Wang, *From Mathematics to Philosophy* (London: Routledge and Kegan Paul, 1974), p. 204. Also Wang, *Logical Journey*, p. 263.

[41] *CW II*, p. 260n20. D.S. Scott, "Measurable Cardinals and Constructible Sets," *Bulletin de l'Académie polonaise des sciences, série des sciences mathématiques, astronomiques, et physiques* 9 (1961), pp. 521–524. S. Ulam, "Zur Masstheorie in der algemeinen Mengenlehre," *Fundamenta Mathematicae* 16 (1930), pp. 140–150.

[42] *CW V*, p. 273; original emphasis.

[43] *CW III*, *1939b, pp. 126–155.

[44] *CW III*, p. 155.

> I therefore would like to point out that the situation in set theory is very different from that in geometry, both from the mathematical and from the epistemological point of view.[45]

Gödel then explains this in terms of weak and strong extensions (see below). In making the comparison with geometry in the Göttingen lecture, he probably did not have the notion of inner model in mind at all, but merely the fact that there are two consistent ways of extending absolute geometry and that it does not make sense to ask which one is the correct one; similarly, he thought at the time, extending ZFC by $V=L$ or by $V\neq L$ are both consistent and it does not make sense to ask which extension is the correct one. (His conviction of the consistency of the axiom stating that nonconstructible sets exist foreshadows in a way the generic sets that Cohen would later use.)

In this lecture Gödel does not explicitly define what he means by "absolutely undecidable," but in his lecture at Brown University in 1940, when referring to the very same result, he defines the related notion of absolute consistency by saying that his consistency proof is absolute in the sense that it is "independent of the particular formal system which we choose for mathematics."[46] By the formal systems that can be chosen he evidently cannot mean just any formal system, as such a system could contain $V\neq L$ as an axiom or an axiom implying it. It is far more likely that he means first of all the systems he also had in mind in the Göttingen lecture the year before, in which he had said that "as is well known, there are different mathematical formalisms, such as the Russellian, the Hilbertian, the formalism of axiomatic set theory, and others";[47] and in addition to those, their extension as suggested by applications of the incompleteness theorem. Indeed, in the Göttingen lecture Gödel went on to mention that "today in fact we know that every mathematical formalism is necessarily incomplete and can be extended by means of new evident axioms. So, strictly speaking, there is no one mathematical formalism at all, but rather only an unsurveyable sequence of ever more comprehensive formalisms." This is a reference to his own incompleteness theorem (as it is this that justifies the adverb in "necessarily incomplete," and thereby his speaking of a "sequence"). The new evident axioms then are the undecidable sentences generated by the proof of the incompleteness theorem (which we can see to be true), in particular, consistency statements, or (more generally) corresponding axioms of infinity (adding new types or levels to the iterative hierarchy). He then says that his consistency proof of *CH* "is applicable to all formalisms hitherto set up, and one can show that it holds unchanged even for the aforementioned extensions by new evident axioms, so that consistency therefore holds in an absolute sense." Because of the reference to the "aforementioned extensions," "absolute consistency" here seems to mean: consistent with ZFC and any series of extensions of it that result from adding statements supplied by the incompleteness theorem. By analogy, "absolutely undecidable" then means: undecidable

---

45 *CW II*, pp. 266–267.
46 *CW III*, p. 184.
47 *CW II*, p. 129.

in ZFC and in any series of extensions of it that result from adding statements seen to be true from the proof of the incompleteness theorem.

That would put Gödel's "absolutely undecidable" statements, which he suggests here includes $V{=}L$, in category 2, but not in category 5. Not without further argument, that is; but that is not to be found in the papers under discussion. Why, then, did he call these statements "absolutely undecidable"? Here we stumble upon a difficulty in Gödel's writings on the theme of undecidability before 1947: besides $V{=}L$ or its negation, he seems to have thought up till then that all axioms to extend ZFC have to be statements seen to be true from the proof of the incompleteness theorem (generally, axioms of infinity), and to have built this into his notion of absolute undecidability. But why would whether a statement is absolutely undecidable or not depend only on ZFC and axioms of infinity? In 1947, Gödel himself suggested that axioms of another type may be needed too. We will come back to this when discussing the paper from that year below.

One is presented with probably this same difficulty by a lecture manuscript which is likely to have been written between 1938 and 1940.[48] Its year is therefore referred to in *CW III* as **193?*, and we will follow this practice. Of interest in this lecture for the present discussion is that Gödel relates his ideas on absolute undecidability explicitly to Hilbert, and that he makes conjectures about the complexity of the simplest absolutely undecidable statements. Instead of ZFC specifically, he here reasons more generally about formal systems on which the only demand is that they can express Diophantine propositions of a specific, simple type. Gödel shows, as he had first done in lectures at Princeton 1934, that the undecidable sentence exhibited in his 1931 paper can be taken to be "almost Diophantine," i.e. of "class A," which is defined as the class of sentences of the form

$$(\forall a_1, \ldots, a_m)(\exists x_1, \ldots, x_n)D$$

where D is a Diophantine equation with natural number coefficients. This is theorem 2 of the manuscript, where theorem 1 asserts the undecidability of class A, in anticipation of the solution of Hilbert's Tenth Problem in the early 1970s due to Matiyasevic, Davis, Robinson and Putnam, which obtains theorem 2 for sentences of class A but with no universal quantifiers. Gödel remarks in the manuscript that the result delineates "the smallest portion of mathematics which cannot be completely mechanized" so far known.[49] This part of the paper, almost the whole, was meant by Gödel as a collection of scattered previous results.[50] But whereas the undecidable statement of class A is of category 1, Gödel suspects that there is a statement of a very similar structure, which is related to *CH*, but behaves very differently:

[48] One suggestion, by John Dawson in *CW III*, p. 163, is that it was prepared with the aim of being presented to the September 1940 International Congress of Mathematicians. That meeting was cancelled due to the outbreak of World War II, and in any case, Gödel never delivered the lecture.
[49] *CW III*, p. 165.
[50] *CW III*, p. 164.

> As to problems with the answer Yes or No, the conviction that they are always decidable, remains untouched by these results [i.e., the existence of undecidable statements in any system that includes class A]. However, I would not leave it unmentioned that apparently there do exist questions of a very similar structure which very likely are really undecidable in the sense which I explained first. The difference in the structure of these problems is only that also variables for real numbers appear in this polynomial. Questions connected with Cantor's continuum hypothesis lead to problems of this type. So far I have not been able to prove their undecidability, but there are considerations which make it highly plausible that they really are undecidable.[51]

What does Gödel mean here by the phrase "really undecidable in the sense which I explained first"? At the beginning of the text, Gödel recalls "Hilbert's famous words that every mathematician is convinced that for any precisely formulated mathematical question a unique answer can be found."[52] Gödel points out that, if this conviction is studied in the context of mathematical logic and proof theory, the incompleteness theorem suffices to refute it even for number theory. However, he adds:

> [I]t is clear that this negative answer may have two different meanings: (1) it may mean that the problem in its original formulation has a negative answer, or (2) it may mean that through the transition from evidence to formalism something was lost.[53] It is easily seen that actually the second is the case, since the number-theoretic questions which are undecidable in a given formalism are always decidable by evident inferences not expressible in the given formalism.[54]

The sense of undecidability that Gödel, as he says at the end of the paper, 'explained first', is the one labelled (1) in this quotation from the beginning of the paper; this means that at the end of the lecture he says that there do seem to be, contrary to Hilbert's conviction, precisely formulated mathematical questions for which no unique answer can be found.[55] Such questions would be of category 5; but to reach such a strong conclusion would seem to be beyond the means available to Gödel then (or later; but see the section on rationalistic optimism, below). Thus, Parsons comments on the closing passage of **193?* that

> It is hard to see what Gödel could have expected to "prove" concerning a statement of the form he describes other than that it is consistent with and independent of the axioms of set theory, say ZF or ZFC, and that this independence would generalize to extensions of ZFC by axioms for inaccessible cardinals in a way that Gödel asserts that his consistency result does.[56]

---

[51] *CW III*, p. 175.
[52] *CW III*, p. 164.
[53] Our footnote: In "Über die Unabhängigkeit der Kontinuumhypothese", *Dialectica* 23 (1969), pp. 66–78, Paul Finsler argues that undecidability of *CH* is a phenomenon that only presents itself in the context of a strict axiomatization, for the "formal continuum." For criticism of this proposal, see e.g. Bernays' remarks on that paper in "Zum Symposium über die Grundlagen der Mathematik", *Dialectica* 25 (1971), pp. 171–195.
[54] *CW III*, p. 164.
[55] Parsons, " Platonism and Mathematical Intuition", p. 67.
[56] P. 50 of C. Parsons, "Platonism and Mathematical Intuition in Kurt Gödel's thought," *The Bulletin of Symbolic Logic* 1(1) (1995), pp. 44–74.

The puzzlement seems to be caused by Gödel's particular and limited view at the time on what absolute undecidability consists in.

We summarize the discussion so far by saying that Gödel seems to have identified for a while categories 2 and 5. In the remainder of this section, we address two questions: What could the polynomials mentioned at the end of the **193?* lecture have been? And did he ever think that $V=L$ is true?

As Gödel says that the polynomials he has in mind are connected to *CH*, one may at first think of equivalents of "Every real is constructible" or of $V=L$. For what other candidates for absolute undecidability could he have had in view? The passage at the end of **193?* bears a close resemblance to one in the Brown lecture of 1940:

> A [every real is constructible] is very likely a really undecidable proposition (quite different from the undecidable proposition which I constructed some years ago and which can always be decided in logics of higher types). This conjectured undecidability of A becomes particularly surprising if you investigate the structure of A in more detail. It then turns out that A is equivalent to a proposition of the following form: $(P)[F(x_1,\ldots,x_k,n_1,\ldots,n_l)=0]$, where F is a polynomial with given integer coefficients and with two kinds of variables x_i, n_i, where the x_i are variables for real numbers and the n_i variables for integers, and where P is a prefix, i.e., a sequence of quantifiers composed of these variables x_i and n_i. I have not yet succeeded in proving that A, and hence this proposition about this polynomial, really is undecidable, but what I can prove owing to the results which I presented in this lecture is of course this: Either this proposition is absolutely undecidable or Cantor's continuum hypothesis is demonstrable (since A implies the continuum hypothesis). But I have not yet been able to determine which one of these two possibilities is realized.[57]

"Every set is constructible" implies "Every real number is constructible", as real numbers are conceived of as particular sets. The converse does not hold, for there exist all kinds of other sets than the reals. However, both imply *CH*, and perhaps that is why Gödel chose to "denote by A or A_n the proposition which says that every real number (and more generally) every set is constructible."[58] On the assumption that the equivalence that Gödel claims indeed exists, we have chosen to gloss A by "every real is constructible"; for by forcing arguments, for no m, n is $V=L$ equivalent to a Π^m_n statement.[59]

"Every real is constructible" does not admit of a Π^1_2-equivalent, by Shoenfield's absoluteness lemma. It is a corollary of this theorem that any Π^1_2 statement is absolute for any transitive model of ZFC that contains all countable ordinals.[60] "Every real is constructible" then cannot be Π^1_2, for there are transitive models of ZFC

---

[57] *CW III*, p. 185.

[58] *CW III*, p. 176.

[59] $V=L$ can be violated "high up" by adding a generic, hence non-contructible, subset to a large regular cardinal, e.g., $\beth_{(\omega+1)}$. The forcing notion used has closure properties which imply that no new subsets of hereditary cardinality $\beth^+_\omega$ are added. Thus all Π^m_n statements are preserved by this forcing for all m and n. This technique was used e.g. in W.B. Easton, "Powers of regular cardinals," *Annals of Mathematical Logic* 1 (1970), pp. 139–178.

[60] E.g., F. Drake, *Set Theory. An Introduction to Large Cardinals* (Amsterdam: North-Holland, 1974), p. 164.

containing all countable ordinals and also non-constructible reals. So neither statement that Gödel denotes by "*A*" is equivalent to a polynomial of the form he has in mind at the end of the paper **193?*; of course it cannot be asked that Gödel had known this in the 1930s. Notice that this particular condition on the form of these polynomials is no longer made in the Brown lecture. This suggests the following possible explanation of the situation: assume that **193?* indeed was written before the Brown lecture.[61] Then it could be that while working on the former, Gödel still suspected that $V=L$ or "every real is constructible' had a Π^1_2-equivalent. In the last line of **193?*, Gödel says about the unspecified polynomials: "So far I have not been able to prove their undecidability, but there are considerations which make it highly plausible that they really are undecidable".[62] In the possible explanation that we suggest, these considerations would involve two stages: first, to establish Π^1_2-equivalents of "Every real is constructible" or of $V=L$, and second, to establish that the latter two "really are undecidable." But in the interval between the two lectures he came to realize (or strongly suspect) that the first stage cannot be completed. Moreover, or as part of this realization, in that interval he had come to see that "every real is constructible" is essentially Π^1_3. The second stage remained, and it is this one that survived in the Brown lecture (and beyond, until Cohen's work).

It has been suggested (e.g., by Martin Davis and by Gregory Moore) that upon introducing $V=L$, at first, Gödel thought that it is true.[63] To be sure, Kreisel reports that "At the time he toyed with the idea that L contained all legitimate definitions of sets";[64] the crucial step to arrive at the identification of V and L would then be to assert that, besides the classical ordinals which are taken as given, no other sets but the legitimately definable (i.e., constructible) exist. As evidence for the suggestion that Gödel indeed identified V and L, Davis and Moore point to a statement that Gödel made when announcing his consistency proof of *CH* in 1938:

> The proposition A [$V=L$] added as a new axiom seems to give a natural completion of the axioms of set theory, in so far as it determines the vague notion of an arbitrary infinite set in a definite way.[65]

Naturality is a fine thing but it does not always extend to plausibility, let alone truth; for would $V=L$ determine the vague notion in the right way? Gödel's formulation leaves this very much open. It qualifies the axiom as natural "in so far as" it sharpens the notion of an arbitrary set "in *a* definite way" (emphasis ours). Even someone who is convinced that $V=L$ is false would agree that it thus sharpens the notion of arbitrary set. Gödel's formulation does not at all exclude that there are other definite ways to determine the vague notion.

---

[61] See *CW III*, p. 163.
[62] *CW III*, p. 175.
[63] *CW III*, p. 163 and *CW II*, p. 158, respectively.
[64] Kreisel, "Gödel's Excursions," p. 158.
[65] *CW II*, p. 27.

An additional suggestion offered by Davis[66] is that Gödel's use of the term "axiom" for $V{=}L$ in his monograph on the consistency of *CH* from 1940[67] is indicative of his holding it true; but Gödel may well have meant to use the term in a formal sense that is not related to truth, as he would do for example on p. 184 of *1947*, in particular when he writes "from an axiom in some sense directly opposite to this [axiom of constructability] the negation of Cantor's conjecture [*CH*] could perhaps be derived."[68]

In 1938, Gödel only mentions the consistency of $V{=}L$ and says nothing about $V{\neq}L$. Did the reason he gives for thinking $V{\neq}L$ is also absolutely consistent occur only later? In any case, during the period that Gödel considered $V{=}L$ as well as its negation absolutely consistent (which period includes the Göttingen lecture from 1939, arguably the lecture **193?*, and the Brown lecture from 1940), he cannot, given his views, reasonably have held $V{=}L$ true. For to hold $V{=}L$ true under those circumstances would be to claim that $V{=}L$ is of category 4, and as we have explained it is obvious that that category should be empty. Note that in Göttingen in 1939, after having named two "interesting consequences" of $V{=}L$(one of which being *CH*), he adds merely that "besides, the consistency of A [$V{=}L$] has a certain interest in and of itself";[69] one would have expected a stronger formulation if he had believed that $V{=}L$ is moreover true.

2.2 1947: . . .but Not for Strong Absolute Undecidability

As $V{=}L$ implies *CH*, any argument against *CH* would also be an argument against $V{=}L$. In 1947, in "What is Cantor's continuum problem?," Gödel adduces a number of reasons why *CH* is probably false. By implication, these are reasons why $V{=}L$ is probably false (and to that extent indicates that $V{=}L$ is not in category 5); indeed, he writes that "not one plausible proposition is known which would imply the continuum hypothesis."[70] The reasons that Gödel presents all consist in a fact and a judgement; the fact being of the form "It has been shown that *CH* has consequence P," and the judgement that P is very implausible or paradoxical.[71] Gödel mentions that these facts were "not known or not existing at Cantor's time." He then gives

---

[66] *CW III*, p. 163.
[67] E.g. *CW II*, p. 81.
[68] *CW II*, p. 184n22. On Gödel's use of the term "axiom" in the Cantor papers, see also H. Wang, *Reflections on Kurt Gödel* (Cambridge: MIT Press, 1988), pp. 205, 294. On p. 205, Wang offers an alternative solution to ours: '[W]hat I see as the main point in this episode is an additional flexibility (besides the allowance for new axioms to be discovered) implicit in Gödel's concept of the axiomatic method: What is thought to be an axiom at one time may later turn out to be a false proposition and, therefore, not really an axiom'. But for the reason we have given, we do not think that Gödel held $V{=}L$ true in the first place.
[69] *CW III*, p. 133.
[70] *CW II*, p. 186.
[71] The judgements that Gödel goes on to make are not universally shared. See Hallett, *Cantorian Set Theory*, p. 111 for some notes of dissent and for further references.

a list of such facts, referring to results published by Luzin in 1914, by Sierpiński between 1924 and 1935 (one of them with Braun), and by Hurewicz in 1932.[72] Given these dates, it is somewhat surprising that Gödel in his lectures in 1939–1940 instead of mentioning them suggests that $V=L$ is not only undecidable in ZFC but "absolutely undecidable." As we have seen, it is not in every case immediately obvious what Gödel meant by that term, its reference seeming to oscillate between categories 2 and 5. But in either case the facts in question might have given him pause: either because they suggest inadequacy of the label "absolutely undecidable" for category 2, or because they suggest that there are considerations leading to a decision of $V=L$ after all, on account of which it would not be in category 5. This strengthens the suspicion, noted above, that there was something missing in his notion of absolute undecidability at that time.

On the other hand, Gödel's willingness in the text **193?* to identify his (particular) notion of absolute undecidability with Hilbert's informal notion (category 5) is at odds with a conviction on which Menger reports. According to Menger's memoir, Gödel in 1939[73] had come to express "more and more emphatically" his

> early conviction that the right axioms of set theory had not yet been discovered [...] He undoubtedly meant that no one had given an adequate basic description of that world of sets in which he believed—a description that would permit us to decide the fundamental problems of cardinality such as Cantor's continuum hypothesis [...] I [Menger] myself never heard from him any indications about where he expected to find such axioms.[74]

It is difficult to see how Gödel could suggest the existence of statements that are absolutely undecidable in Hilbert's original sense if he at the same time thought that axioms were still missing. The "early conviction" Menger had described in somewhat more detail earlier on: "In 1933 he already repeatedly stressed that *the right* (*die rechten*, sometimes he said *die richtigen*) *axioms of set theory had not yet been found*."[75]

3 1947: *CH*, Conceptual Incompleteness and Realism

The question raised by Menger's memoir is perhaps not unanswerable, but at present we have no suggestion to make. The fact remains that, even if Menger is correct about what Gödel told him in 1939, in the lectures of 1939–1940 (the Göttingen lecture in 1939 took place on December the 15th, so after the stay at Notre Dame that Menger reports on, which lasted from January till June), Gödel certainly breathed no word about this conviction that fundamental axioms were still missing from set theory. In 1947, however, he came to communicate it publicly. From a philosophical point of view, the particular form this suggestion takes is of a much broader

---

[72] *CW II*, p. 185–186.
[73] Menger's '1938' on p. 220 of his *Reminiscences* must be a slip of the pen.
[74] Menger, *Reminiscences*, p. 222.
[75] Menger, *Reminiscences*, p. 210; original emphasis.

importance (because it pertains directly to the very foundations) than a decision of the specific problem of *CH* (by perhaps known means) would be:

> As for the continuum problem, there is little hope of solving it by means of those axioms of infinity which can be set up on the basis of principles known today (the above-mentioned proof for the undisprovability of the continuum hypothesis, e.g., goes through for all of them without any change). But probably there exist others based on hitherto unknown principles; also there may exist, besides the ordinary axioms, the axioms of infinity and the axioms mentioned in footnote 17 [axioms on higher-order properties of sets], other (hitherto unknown) axioms of set theory which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts.[76]

Most of the subsequent attention of set theorists to this passage seems to have gone into "axioms of infinity based on hitherto unknown principles." Yet the most important difference with the 1939–1940 lectures is that Gödel here has come to consider the need for new axioms whose introduction is not suggested by the incompleteness theorem but rather by conceptual analysis (this emphasizes that incompleteness cannot be considered merely an artifact of formalization). It might of course happen that justifications from the concept of set will also be found for the large cardinals based on new principles. In 1966, Gödel pointed out that so far this had not happened.[77] As Charles Parsons remarked on the lecture **193?*, "There seems to be a clear conflict with the position of 1947; it's hard to believe that at the earlier time he thought that exploration of the concept of set would yield new axioms that would decide them [i.e. the statements Gödel in **193?* suspected to be "really undecidable"]."[78] (In fact, the large cardinal program to decide *CH* has so far not provided a decisive solution.) In the 1947 paper, Gödel announced the idea of conceptual analysis a few pages before the quotation just given, as follows:

> This scarcity of results, even as to the most fundamental questions in this field, may be due to some extent to purely mathematical difficulties; it seems, however [. . .] that there are also deeper reasons behind it and that a complete solution of these problems can be obtained only by a more profound analysis (than mathematics is accustomed to give) of the meanings of the terms occurring in them (such as "set," "one-to-one correspondence," etc.) and of the axioms underlying their use.[79]

The suggestion, then, is that the usual systems of set theory such as ZFC, as well as being formally incomplete as shown in the incompleteness theorems, are also incomplete in another, more basic sense; they may be called "conceptually incomplete."[80] It is not at all impossible that Gödel's newly found interest in the analysis of concepts was related to his study of Leibniz, but at present we cannot be

---

[76] *CW II*, p. 182.
[77] *CW II*, p. 260n20.
[78] Parsons, *Platonism and mathematical intuition*, p. 50.
[79] *CW II*, p. 179.
[80] This is to be distinguished from yet another notion of incompleteness, described in R. Carnap, "Die Antinomien und die Unvollständigkeit der Mathematik," *Monatshefte für Mathematik und Physik* 41 (1934), pp. 263–284: certain metamathematical notions can be defined in a given system, but others (notably, truth) cannot. The difference with conceptual incompleteness is that such

more specific. There certainly is a strong Leibnizian flavour to an item in notebook XIV which is related to the "concepts underlying logic and mathematics" that he mentioned in the quotation before the last one:[81]

> The fundamental philosophical concept is cause [. . .] Perhaps the other Kantian categories (that is, the logical [categories], including necessity) can be defined in terms of causality, and the logical (set-theoretical) axioms can be derived from the axioms of causality. (Property = cause of the difference of things).[82]

Perhaps it is by such metaphysical derivations that Gödel hoped to clarify a fundamental underdeterminedness of the concept of set by ZFC that he mentions in 1947: one can take as a model for ZFC either his hierarchy L, or the class of arbitrary multitudes irrespective of whether or not they are constructible or in some other sense definable. But presumably it would be an essential property of sets if they are definable. "This characteristic of sets, however, is neither formulated explicitly nor contained implicitly in the accepted axioms of set theory," Gödel comments, and to that extent ZFC is, given these two very different types of models it admits, not sharp enough an axiomatization.[83]

Gödel takes this view because he is a realist, meaning that he is "someone who believes [the axioms of set theory] describes some well-determined reality," in which, in particular, "Cantor's conjecture must be either true or false."[84] Kreisel aptly remarks that the constructible may also be taken to constitute "some well-determined reality";[85] but there is a consideration that would limit the use of that observation as an independent argument for holding that $V{=}L$ is true. Namely, if one holds that mathematical reality should admit of a conceptual description that is entirely self-coherent, this certainly counts against $V{=}L$:

> [The constructibility hypothesis] is not a conceptually pure proposition because it allows ordinal numbers definable only by impredicative definitions or not definable at all, but proceeds to reject all further uses of impredicative definitions.[86]

(Borrowing a term Gödel once used to describe Hilbert's formalism,[87] from a philosophical point of view one may describe L as "a curious hermaphroditic thing.") After his remark on mathematical reality, Gödel concludes about CH that

> its undecidability from the axioms as known today can only mean that these axioms do not contain a complete description of this reality: and such a belief is by no means chimeri-

indefinability results do not arise because of insufficient analysis of the concepts used in the sytem. Gödel's review of Carnap's paper is in *CW I*, p. 389.

[81] From the editorial note on p. 428 of *CW II*, we gather that notebook XIV was used from April 1946 till 1955.

[82] *CW III*, pp. 433–435.

[83] *CW II*, p. 183.

[84] *CW II*, p. 181.

[85] Kreisel, *Review*, p. 607.

[86] Wang, *From Mathematics to Philosophy*, p. 196.

[87] *CW III*, p. 379.

> cal, since it is possible to point out ways in which a decision of the question, even if it is undecidable from the axioms in their present form, might nevertheless be obtained.[88]

(It is not obvious that only realists should find this sufficient reason to look for new axioms.) He then describes two such ways, the one being that of conceptual analysis and the other that of inductive arguments. We will discuss them below, taking as our point of departure the version from 1964 (in which conceptual analysis is tied to a specific notion of intuition).

An interesting example of the possibility of such conceptual advancement Gödel gives in both versions is that of the inaccessible and the Mahlo cardinals. This example is based on the iterative conception of set:

> This 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, has never led to any anitomy whatsoever; that is, the perfectly "naïve" and uncritical working with this concept of set has so far proved completely self-consistent.[89]

Indeed, to Wang he later said that the iterative concept is "simply the correct" concept of set. [90] It is this concept that he has in mind when he writes that

> 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.[91]

The idea is that, as soon as one has determined exact ways of forming sets, all the sets obtained by these specific means can be collected to form a set. If one thinks of the ZFC axioms as a list of such exact means and then applies this idea, one is led to inaccessible cardinals, and from there to the even larger Mahlo cardinals.[92] Both give natural extensions:

> [T]hese axioms show clearly, not only that the axiomatic system of set theory as known today 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.[93]

One might have thought that the existence of inaccessibles requires a separate assumption, involving some form of maximality, to be adjoined to the pure concept of set; but this is not the case.[94]

---

88 *CW II*, p. 181.
89 *CW II*, p. 180.
90 Wang, *Logical Journey*, p. 238.
91 *CW II*, p. 181.
92 κ is Mahlo if $\{lambda < \kappa | \lambda$ is inaccessible$\}$ is stationary.
93 *CW II*, p. 182.
94 In the 1964 version, Gödel amended this sentence to read "but also that it can be supplemented without arbitrariness by new axioms which only unfold the content of the concept of set explained above," which perhaps reflects the study of phenomenology that he had begun in between the two versions of the Cantor paper; in the German phenomenological argot, 'Entfaltung', or the more or less synonymous 'Auslegung', 'Explikation', 'Auseinanderlegung', 'Explizierung' are very com-

4 Abstract Considerations About Absolute Undecidability

4.1 1944, 1946: Absolute Provability

As was only to be expected, the different developments in Gödel's thought concerning these topics did not dovetail neatly but overlapped. We take a small step back in time. Only a few years after writing the manuscript **193?*, which leaves open the possibility that there exist strongly absolutely undecidable sentences, Gödel came to think that, on the contrary, category 5 is empty. In 1946, in his remarks before the Princeton bicentennial conference on problems in mathematics,[95] Gödel commented briefly on a notion of absolute demonstrability (absolute in the sense of not depending on the formalism chosen). Such a concept of demonstrability could of course not be entirely formalizable (because of his own incompleteness theorem), but Gödel does not exclude that a concept of an appropriately different character can be found which would entail the decidability of every set-theoretic proposition:

> It is not impossible that for such a concept of demonstrability some completeness theorem would hold which would say that every proposition expressible in set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets.[96]

As we saw above, by 1947 Gödel thought that axioms of infinity need not be sufficient and that axioms of a different kind may also be required (as in the context of this quotation Gödel only speaks of axioms of infinity, we take it that the one time he uses "largeness" he does not also have in mind the width of the hierarchy; also note that largeness may well involve more than just cardinality). What is in any case striking is the very suggestion here that a notion of absolute provability (for set theory) is possible and moreover within reach. The philosophical attitude required to make such a remark with some confidence may well have been instilled or reinforced in Gödel by his study of Leibniz, for the remark may be considered as a further development of the claim Gödel had made at the end of the Russell paper, two years earlier:

> Leibniz did not in his writings about the *Characteristica universalis* speak of a utopian project; if we are to believe his words he had developed this calculus of reasoning to a large extent [...] He went even so far as to estimate the time which would be necessary for his calculus to be developed by a few select scientists to such an extent "that humanity would have a new kind of an instrument increasing the powers of reason far more than any optical instrument has ever aided the power of vision." [...] Furthermore, he said repeatedly that, even in the rudimentary state to which he had developed the theory himself, it was responsible for all his mathematical discoveries.[97]

---

mon terms. Note that Gödel had spoken of "explicating the content of the general concept of set" already in version III of his paper on Carnap, *CW III*, p. 353n43.

[95] *CW II*, p. 150–153.

[96] *CW II*, p. 151.

[97] *CW II*, p. 140–141.

What Gödel says here is amplified by a remark he is recorded to have made in 1948. Wang reports on a note by Carnap on a conversation with Gödel on March 3 of that year, according to which Gödel thought that Leibniz apparently had obtained a decision procedure for mathematics.[98] Gödel also said that, while the system cannot be completely specific (again, because of his own incompleteness theorem), it may still give sufficient indications as to what is to be done.[99]

At the same conference in Princeton in 1946, Tarski also spoke on decision problems.[100] He makes the distinction (for number theory) between undecidable statements of category 1 and 2; as for problems in set theory, he mentions Gödel's recent work on the continuum hypothesis and expresses a belief that certain problems of set theory may be independent (as we saw above, he had done the same in 1929, when the actual situation in set theory had been less clear). But unlike Gödel, he does not touch on the problem whether category 5 is empty or not.[101]

A remark Church made in the discussion at the Princeton conference should be noted as well.[102] Zermelo had in 1932 proposed a theory of infinite proofs and had hoped that all true mathematical propositions were provable in this extended sense. Church objected to proposals of this kind (as reported in the minutes) that "while such systems might have considerable interest of one kind or another, they could not properly be considered *logics*, insofar as logics explicate the notion of *proof*. For what we mean by a proof is something which carries finality of conviction to any one who admits the assumptions (axioms and rules) on which the proof is based; and this requires that there be an effective (finitary, recursive) syntactical test of the validity of proposed proofs."

4.2 1951: Strong Absolute Undecidability as an Abstract Possibility

In 1951 Gödel returns to absolute undecidability. In what has become known as the Gibbs lecture, he defines absolute undecidability to mean "undecidable, not just

98 Wang, *Reflections*, p. 173.

99 Wang, *Reflections*, p. 174.

100 Tarski's paper (and some additional material) has been transcibed, edited and introduced by H. Sinaceur, "Address at the Princeton University Bicentennial Conference on Problems of Mathematics (December 17–19, 1946), by Alfred Tarski," *The Bulletin of Symbolic Logic* 6(1) (2000), pp. 1–44. Without hazarding an explanation, we note the following difference between Gödel's talk and a report on a discussion session at the same meeting. In the talk, Gödel says that '[I]t has some plausibility that all things conceivable by us are denumerable, even if you disregard the question of expressibility in some language' (*CW II*, p. 152); in the report, Gödel is said to propose to allow uncountably many primitive notions in a formal system, and is then quoted as saying "I do not feel sure that the set of all things of which we can think is denumerable" (Sinaceur, "Address," p. 37). Technically, one can of course admit this extension of the notion of formal system, but the quoted remark evidently adds an epistemological concern.

101 Related material is in Tarski's letter to Gödel of Dec. 10, 1946 (one week before the conference), *CW V*, pp. 271–273, and in Feferman's remarks on p. 264.

102 Sinaceur, "Address," pp. 33–34.

within some particular axiomatic system, but by *any* mathematical proof the human mind can conceive."[103] (As we already had occasion to recall, in version III of the Carnap paper Gödel characterized the notion of proof in "its original 'contensive' meaning" as "a sequence of thoughts convincing a sound mind".[104]) This time there is no ambiguity, and he clearly means strong absolute undecidability. In particular he considers the possibility that among the absolutely undecidable sentences in this sense, if there are such, will occur standard Diophantine sentences of type Π^0_2. He then goes on to establish his "disjunctive theorem":

> Either mathematics is incompletable in this sense, that its evident axioms can never be comprised in a finite rule, that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable Diophantine problems of the type specified (where the case that both terms of the disjunction are true is not excluded, so that there are, strictly speaking, three alternatives).[105]

Truth of the first disjunct of course would not mean that category 4 in our classification of undecidable sentences is non-empty after all. It is rather based on the fact that the capacity to see the consistency of every consistent finite formal system is not a capacity that a finite machine can have; so if the human mind indeed has that capacity, it is not a finite machine. Its powers would moreover surpass that of any finite machine "infinitely," because for any finite machine there exist infinitely many others of which that machine cannot establish their consistency.

As Gödel adds, he means the disjunction to be inclusive: thereby the possibility that category 5 is non-empty is, in effect, explicitly left open. That Gödel considers it plausible that it is not empty may be inferred from his characterization, later on in the Gibbs lecture, of his platonistic view as "the view that mathematics describes a non-sensual reality, which exists independently both of the acts and [of] the dispositions of the human mind and is only perceived, and probably perceived very incompletely, by the human mind."[106] An alternative explanation would be that perception only allows us to establish basics such as the axioms, and that for more complicated cases perceptions are not available and we have to resort to logic. Below we will make some remarks about Gödel's realist views on mathematics; here we would like to emphasize a point made by Charles Parsons that the existence of (strongly) absolutely undecidable propositions would in itself not be incompatible with realism.[107]

---

[103] *CW III*, p. 310.

[104] *CW III*, p. 341 n. 20.

[105] *CW III*, p. 310. Are Diophantine sentences of type Π^0_2 elementary? From the point of view of strong and interesting theories like $I\Delta_0 + \Omega_1$ the Π^0_2-theory of the natural numbers looks very powerful; highly intractable.

[106] *CW III*, p. 323.

[107] Parsons, "Platonism and Mathematical Intuition', p. 52; see also note 16 on the same page for some remarks on Gödel's realism in the light of Dummett's conception of realism.

4.3 Phenomenology and Rationalistic Optimism

One consequence of the disjunctive theorem is this: If the mind is a finite machine, then there are absolutely undecidable Diophantine problems. So one might try to settle the issue by attempting to establish that the mind indeed is a finite machine.

However, that was clearly not what Gödel had in mind, given the views he expressed on Leibniz in the 1940s (see above), and, consistent with Leibniz' position as the (grand)father of German Idealism, Gödel's philosophical development in the direction of idealistic philosophy; in particular, from 1959 on, to Husserl's transcendental idealism, which became the general framework for his general philosophical endeavours and for the grounding of his mathematical realism in particular.[108] One of Gödel's aims was to use phenomenology to clarify our understanding of the mind as well as of the ontology of mathematics to such an extent that it would be established that the mind is not a finite machine, and that there are no absolutely unsolvable problems.[109] In a draft letter from (June?) 1963 from Gödel to TIME Inc., regarding the upcoming publication *Mathematics* in the Life Science Library, he connects his phenomenological program to his famous "disjunctive conclusion" that either the human mind infinitely surpasses the powers of any finite machine, or there exist absolutely unsolvable Diophantine problems.[110] In that draft letter, he mentions the disjunction again, with the disjuncts in reverse order, and then comments:

> I believe, on ph[ilosophical] grounds, that the sec[ond] alternative is more probable & hope to make this evident by a syst[ematic] developm[ent] & verification of my phil[osophical] views. This dev[elopment] & ver[ification] constitutes the primary obj[ect][111] of my present work.[112]

And another version of that passage reads

> I conj[ecture] that the sec[ond] altern[ative] is true & perhaps can be verified by a phenomenol[ogical] investigat[ion] of the processes of reasoning.[113]

---

[108] In the discussion at the Princeton conference in 1946, Tarski noticed that idealistic philosophy is congenial to some of Gödel's views on mathematics (Sinaceur, "Address," p. 34). Menger reports that Gödel in the thirties "studied a great deal of philosophy including post-Kantian German idealist metaphysics" (Menger, *Reminiscences*, p. 209). See M. van Atten and J. Kennedy, "On the Philosophical Development of Kurt Gödel," *The Bulletin of Symbolic Logic* 9(4) (2003), pp. 425–476, for further discussion of Gödel's interest in idealism.

[109] The remainder of this paragraph, until and including the two quotations from Gödel, is extracted from p. 460 of Van Atten and Kennedy, "On the Philosophical Development." For the full text of the draft letter, as well as a draft letter from Gödel to Tillich on the same subject, see M. van Atten, "Two draft letters from Gödel on self-knowledge of reason", *Philosophia Mathematica* 14(2), 2006, pp. 255–261.

[110] *CW III*, p. 310.

[111] After "obj[ect]," Gödel at first wrote "matter" but then crossed it out.

[112] Gödel Nachla (GN), Firestone Library, Princeton, item 020514.7.

[113] Gödel first wrote "thin[king]," and then crossed it out.

A sign of Gödel's optimism at the time is that he saw to it that in the TIME book itself, which appeared in 1963, it was reported that

> "Either mathematics is too big for the human mind,' he says, "or the human mind is more than a machine." He hopes to prove the latter.[114]

In our discussion of the paper from 1964, we will make some comments on the importance of phenomenology for Gödel's realism. In later remarks on minds and machines, Gödel brings into play what he calls "the rationalistic attitude," in connection to which he mentions the name of Hilbert but which also takes up again the Leibnizian theme at the end of the Russell paper. In the 1970s, Gödel said to Wang:

> Our incompleteness theorem makes it likely that the mind is not mechanical, or else the mind cannot understand its own mechanism. If our result is taken together with the rationalistic attitude that Hilbert had *and which was not refuted by our results*, then (we can infer) the sharp result that the mind is not mechanical. This is so, because, if the mind were a machine, there would, contrary to this rationalistic attitude, exist number-theoretic questions undecidable for the human mind.[115]

In 1972 he went into a little more detail and gave the basic ideas of two arguments, which ideas were then published in Wang's *From Mathematics to Philosophy*:

> If it were true [that there exist number theoretical questions undecidable for the human mind] it would mean that human reason is utterly irrational by asking questions it cannot answer, while asserting emphatically that only reason can answer them. Human reason would then be very imperfect and, in some sense, even inconsistent, in glaring contradiction to the fact that those parts of mathematics which have been systematically and completely developed (such as, e.g. the theory of 1st and 2nd degree Diophantine equations, the latter with two unknowns) show an amazing degree of beauty and perfection. In these fields, by entirely unexpected laws and procedures (such as the quadratic law of reciprocity, the Euclidean algorithm, the development into continued fractions, etc.), means are provided not only for solving all relevant problems, but also solving them in a most beautiful and perfectly feasible manner (e.g. due to the existence of simple expressions yielding *all* solutions). These facts seem to justify what may be called "rationalistic optimism."[116]

The first argument is a deduction from the essence of reason. If one wishes to attempt such an argument, it would be natural to do so in the context of phenomenology, and this is what Gödel will have had in mind. It would go together well with his intention (see above) to apply phenomenology to establish that the mind infinitely surpasses any finite machine. Similarly, Gödel's claim that "In principle, we can know all of mathematics. It is given to us in its entirety and does not change—unlike the Milky Way."[117] is probably more easily interpreted in the context of Husserl's transcendental idealism than in others.

---

[114] D. Bergamini and the editors of Life, *Mathematics* (New York: Time, 1963) p. 53.
[115] Our emphasis; Wang, *Logical Journey*, pp. 186–187.
[116] Wang, *From Mathematics to Philosophy*, pp. 324–325. See also Wang, *Logical Journey*, p. 316, item 3, and p. 317. For a connection to Kant here, see Boolos' remark in *CW III*, p. 294.
[117] Wang, *Logical Journey*, p. 151, 4.4.18.

The second argument is a projection from very specific, highly successful theories. This is a wholly different kind of argument. It is not in obvious contradiction with phenomenological principles but it would take further work to see exactly how it fits in with them. We notice that, as he would do in the 1964 version of the Cantor paper, Gödel here gives two types of argument for a strong conviction: one based on intuition (here, of essences) and one from success (of reason in a particular area). There is a comment by Gödel that is related to this second argument and that contains a reflection on the fact that Hilbert and he shared the conviction of the decidability of all mathematics:

> We have the complete solutions of linear differential equations and second-degree Diophantine equations. We have here something extremely unusual happening to small sample; in such cases the weight of the sample is far greater than its size. The a priori probability of arriving at such complete solutions is so small that we are entitled to generalize to the large conclusion, that things are made to be completely solved. Hilbert, in his program of finitary consistency proofs of strong systems, generalized in too specialized a fashion.[118]

(We have not investigated to what extent this view on what Hilbert did is historically accurate.) We will now see how Gödel in 1964 brought the strongly rationalist position which he is likely to have held from early on but took many years to articulate to bear on *CH*.

5 1964: How to Find New Axioms and Decide *CH*

5.1 The Meaningfulness of the Question

In the supplement to the 1964 edition of the Cantor paper, Gödel gives two criteria for determining whether a statement that is independent of ZFC gives rise to a decision problem that is meaningful. The first is a mathematical criterion. It is in a sense a result in meaning analysis; on the other hand, at the time Gödel could not demonstrate but only make plausible that *CH* satisfies it. The second is a philosophical criterion. If one accepts the philosophical position that motivates that criterion, then *CH* certainly satisfies it.

The mathematical criterion is based on a distinction between different kinds of extensions of axiomatic systems. Consider the parallel axiom in geometry. Both it and its negation are independent of the first four axioms (absolute geometry), which can thus be extended either way, but for both extensions one can find models in the unextended (Euclidean) system. But then the question of the truth (simpliciter) of the parallel postulate "became meaningless for the mathematician."[119] Rather, geometry bifurcates at the parallel axiom. Gödel speaks of "weak extensions." Something similar holds for questions about extensions of the real field by the addition or non-addition of complex numbers.

---

118 Wang, *Logical Journey*, p. 317, 9.4.21.
119 *CW II*, p. 267.

Gödel then considers extensions that are stronger, in the sense that they are not weak extensions and that moreover they also have consequences outside their own domain. Gödel gives the example of inaccessible cardinals. In ZFC we can define a model of ZFC + the statement "there are no inaccessible cardinals" as follows:

Case 1. Suppose there are no inaccessible cardinals. Then V can be taken to be the desired model.
Case 2. Suppose there is an inaccessible κ. Take the least such, and call it λ. Cut the universe at V_λ and take everything below for the desired model.

Note that no axiom beyond ZFC has been invoked in the construction of the model, so the statement "there are no inaccessible cardinals" is a weak extension of ZFC. It does not result in new theorems about integers. On the other hand, it is easy to see that in ZFC one cannot establish a model of ZFC + "there exists an inaccessible cardinal" this way. The latter therefore is not a weak extension, and, moreover, new theorems about integers follow from it. Hence it is an extension in a stronger sense. Gödel's mathematical criterion is then, that the question as to the truth of an independent statement is meaningful if either it or its negation (and presumably not both) would be a stronger extension of this type. Applied to *CH*, Gödel notes that models of ZF+*CH* can be obtained by an inner model construction; also *CH* is "sterile for number theory,"[120] i.e. *CH* implies no new theorems about the integers. Therefore *CH* is a weak extension of ZF. Models of ZFC+¬*CH* cannot be thus obtained (Shepherdson's result, see above) and assuming that alternatives to *CH* may have consequences outside their domain, the question whether *CH* is true or false remains meaningful even though it is independent of ZFC. That would show a difference between the parallel postulate and *CH*. As it turns out, a simple forcing argument demonstrates that the negation of *CH* is also sterile for number theory, as noted by Gödel in the postscript to the paper. An asymmetry between the parallel postulate and the *CH* lying in a somewhat opposite direction has been pointed out by Kreisel: namely, in Hilbert's second order axiomatization the parallel postulate is still independent, whereas second order *CH* is decidable.[121] This demonstrates, in Kreisel's view, the significance of the first order/second order distinction. In the section on the notion of success below, we will return to this mathematical criterion.[122]

The philosophical criterion is based on Gödel's realism. In the 1947 version, he had already given the argument that *CH* has a fixed truth value because it is a proposition about a well-determined reality. In our discussion of *1947* we have alluded to this passage but did not quote it. In 1964 he repeats this passage (with

---

[120] *CW II*, p. 267.
[121] See G. Kreisel. "Informal rigour and completeness proofs". In I. Lakatos, editor, *Problems in the philosophy of mathematics* (Amsterdam: North-Holland, 1967), pp. 138–186.
[122] For more about Gödel's criterion, see also Kennedy's "Gödel and Meaning," in preparation.

some changes in the formulation, which we have argued elsewhere reflect his study of Husserl):[123]

> For if the meaning of the primitive terms of set theory as explained on page 262 and in footnote 14 are accepted as sound, it follows that the set-theoretical concepts and theorems describe some well-determined reality, in which 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; and such a belief is by no means chimerical, since it is possible to point out ways in which a decision of the question, even if it is undecidable from the axioms in their present form, might nevertheless be obtained.[124]

Robert Tragesser, in his book *Phenomenology and Logic*, explains that this is "the extremely crucial statement" in Gödel's considerations as to how the continuum problem might be solved:

> What is so important in this statement is the tie it makes between our right to say that S [i.e., the domain of set theory] is a well-determined reality (in which, say, *CH* is decided) and the discoverability of promising ways in which open problems (e.g., *CH*) about the domain could be decided. Gödel spends the remainder of the article presenting possible paths to a decision about *CH*. As long as we can find such paths, S will seem to be the well-determined reality we initially took it to be.[125]

The tie that Gödel makes here reflects his adoption of Husserl's transcendental idealism from 1959 onward. For one aspect of Husserl's later philosophy that was of particular importance to Gödel was the way it analysed the relations (1) between the existence of (concrete as well as abstract) objects and consciousness and (2) between consciousness and reason. Briefly, the basic principle of transcendental idealism is that the objects that can be said to exist are exactly those that are in principle accessible to a consciousness that acts in accordance with the evidence it obtains for those objects. To act in that way is precisely what rationality in the most pregnant sense consists in. Thus, Gödel's realism, after having received its foundation in transcendental idealism, and his rationalism, are intimately connected. One can even say that they are two sides of the same coin.[126]

Tragesser continues:

> Gödel may be viewed as giving an analysis of the elements of the prehension of S and, on the foundations of that analysis, showing how *CH* could possibily be decided. Such analysis, because it reflects faithfully upon, and describes, the elements of an act of consciousness (a prehension, in this case), is *phenomenological analysis*. We can see here the critical importance of such analysis, viz., that it provides possible paths to reasons better than arbitrary for holding something to be true of a considered object or objective domain.[127]

---

[123] See Van Atten and Kennedy, *On the Philosophical Development*, Section 6.3.

[124] *CW II*, p. 260. The two references Gödel makes are to a description of the iterative concept of set and a remark that the existence of a set does not presuppose that it is definable in a finite number of words.

[125] R. Tragesser, *Phenomenology and Logic* (Ithaca: Cornell University Press, 1977), p. 22.

[126] For this paragraph, see Van Atten and Kennedy, *On the Philosophical Development*.

[127] Original emphasis; Tragesser, *Phenomenology and Logic*, pp. 23–24.

Tragesser defined "prehension" as the "imperfect or incomplete "grasp" of a purportedly objective state of affairs, where it is somehow known that the state of affairs is imperfectly or incompletely given."[128] ZFC expresses a prehension of the universe, in the sense that ZFC is not a complete axiomatization in the two senses Gödel gives to the word "incomplete." Gödel suggests that on the basis of this prehension, in other words, starting from the "incomplete description" we have gotten of the set-theoretic universe so far, it is possible to proceed in such a way as to decide *CH*. In the paper, he proposes two "truth criteria"[129] for evaluating candidate axioms extending ZFC.

5.2 Two Truth Criteria for New Axioms

5.2.1 Intuition

The strong notion of intuition invoked in the 1964 version of the Cantor paper is one of the most conspicuous differences with the text from 1947. In the larger context of Gödel's philosophical development, it is natural that it should have appeared, given Gödel's intensive study of (and enthusiasm for) Husserl's phenomenology since 1959.[130] The key passage in the 1964 paper is this one:

> But, despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception.[131]

Much has been written on the interpretation of this passage, and we refer the reader to Parsons,[132] Tieszen,[133] Tragesser,[134] Van Atten and Kennedy,[135] and, for background, Husserl, whose sixth *Logical Investigation*[136] Gödel recommended to logicians in the 1960s.[137]

---

[128] Tragesser, *Phenomenology and Logic*, 17.
[129] *CW II*, p. 269.
[130] It is not excluded that in 1947 Gödel was playing with ideas prefiguring the notion of intuition known from the 1964 paper. Charles Parsons, ("Platonism and mathematical intuition," p. 57) has drawn attention to allusions to perception of concepts in the paper on Russell from 1944 and in the Gibbs lecture from 1951. We agree with Parsons that "it is reasonable to conjecture that although [Gödel at that time] was not yet ready to defend his notion of intuition he already had some such conception in mind" (p. 57).
[131] *CW II*, p. 268.
[132] "Platonism and Mathematical Intuition."
[133] "Gödel and the Intuition of Concepts," *Synthese* 133(3) (2002), pp. 363-391.
[134] *Phenomenology and Logic.*
[135] "On the Philosophical Development," Section 6.2.
[136] E. Husserl, *Logische Untersuchungen. Zweiter Band, 2. Teil* Husserliana vol. XIX/2 (Den Haag: Martinus Nijhoff, 1984).
[137] Wang, *Logical Journey*, p. 80.

The comparison of (abstract) intuition to (sense) perception in this passage shows that Gödel means intuition in a technical sense, and as such it is just as decisive as the intuitionists intend it to be. He is not talking about intuition as (merely) a psychological fact here. Still, Gödel allows for mistakes even in intuitions, but that is because intuition is not an all-or-nothing affair. It comes in degrees.[138] And ultimately, existence remains tied to (ideal) intuition, by the basic principle of transcendental idealism.

The work that this intuition is meant to do with respect to the continuum problem is to give a well-defined meaning to the question and indeed to decide it. In 1964, Gödel writes, "That new mathematical intuitions leading to a decision of such problems as Cantor's continuum hypothesis are perfectly possible was pointed out earlier (pp. 264–265)."[139]

Gödel was aware that his talk of an objective realm of transfinite set theory and of a faculty of intuition that has access to it would probably not be well received. In fact, he had already feared that the 1947 version, which did not even contain the strong views on mathematical intuition yet, would be subjected to positivistic attacks by Benacerraf and Putnam in the introduction of their planned anthology which was to contain it.[140] Gödel not only overcame these fears but he went ahead to include views even more opposed to positivism in the revision of the 1947 paper that he went on to prepare for the occasion. But perhaps it was a residual fear that motivated him to propose also an argument from metaphysically less contentious premises that should lead to the desired conclusion that *CH* can be decided.

This argument takes the existence of intuition not as an epistemological but as just a psychological fact (where the former sense does, and the latter does not, imply access to objects in reality):

> However, the question of the objective existence of the objects of mathematical intuition [...] is not decisive for the problem under discussion here [...] The mere psychological fact of the existence of an intuition which is 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.[141]

By an appeal to psychology, Gödel suggests that meaningfulness and decidability may be securable without resorting to realism. In 1975, Hao Wang characterized this statement "as asserting the possibility of recognizing meaningfulness without realism"; Gödel agreed, for he suggested that Wang would report: "He [i.e., Gödel] himself suggests an alternative to realism as ground for believing that undecided propositions in set theory are either true or false."[142] Given Gödel's avowals of

---

[138] See also Husserl, e.g. Section 24 of Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Erstes Buch Husserliana vol. III/1 (Den Haag: Martinus Nijhoff, 1976).

[139] *CW II*, p. 268. Gödel's page references are to the original publication, which correspond to pp. 260–261 in *CW II*.

[140] *CW II*, p. 166

[141] *CW II*, pp. 268–9.

[142] Wang, *Logical Journey*, p. 243.

realism in this paper[143] and elsewhere, it is clear that he does not actually embrace this alternative. But suggesting this alternative serves a dialectical purpose to him, that of indicating that even from alternative points of view, his idea (here, of meaningfulness of the question) is the correct one.[144]

The argument from the psychological fact however is not particularly strong, for a reason that Gödel himself had indicated before and that he, surprisingly, left out of consideration both in the 1964 paper itself and, apparently, in his discussion with Wang of the passage that we just quoted. This reason is stated in a footnote to version III of the paper on Carnap, **1953/9-III*:

> the existence, as a psychological fact, of an intuition covering the axioms of classical mathematics can hardly be doubted, not even by adherents of the Brouwerian school, except that the latter will explain this psychological fact by the circumstance that we are all subject to the same kind of errors if we are not sufficiently careful in our thinking.[145]

This seems to give a more complete view of the situation but also a somewhat more complicated one, for should the intuitionists be right, one can doubt whether an intuition that is brought about by insufficiently careful thinking can itself function as the basis for a sufficiently careful judgement (a judgement that, if it can be made, from the intuitionist's point of view will be purely hypothetical or "as if"). So for the "psychological fact" to have the force that Gödel takes it to have, the suggested intuitionistic interpretation has to be shown wrong first. It would be crucial to do this at some point, for Gödel's whole approach of an appeal to intuition to clarify and (to find axioms that) decide *CH* stands or falls with his being able to come up with a notion of intuition that is relevantly different from the intuitionist's.

Of course Gödel does not at all doubt that the intuitionist is indeed wrong, as is clear from a passage that at the same time brings out the urgency of finding an argument to that effect:

> First of all there is Brouwer's intuitionism, which is utterly destructive in its results. The whole theory of $\aleph$'s greater than $\aleph_1$ is rejected as meaningless [...] However, this negative attitude toward Cantor's set theory, and toward classical mathematics, of which it is a natural generalization, is by no means a necessary outcome of a closer examination of their foundations, but only the result of a certain philosophical conception of the nature of mathematics, which admits mathematical objects only to the extent to which they are interpretable as our own constructions or, at least, can be completely given in mathematical intuition. For someone who considers mathematical objects to exist independently of our constructions and of our having an intuition of them individually, and who requires only that the general mathematical concepts must be sufficiently clear for us to be able to recognize their soundness and the truth of the axioms concerning them, there exists, I believe,

---

143 Wang (*Reflections*, p. 292) writes that according to Gödel, certain sections in this paper, together with his contribution to Wang's book *From Mathematics to Philosophy*, "constitute the principal statement of his mathematical realism."

144 For other examples of this maneuver, see *CW III*, p. 345 n45, p. 356 lines 3–5, p. 361 line -14; also *Logical Journey*, p. 239, 7.4.7.

145 *CW III*, p. 338 n12. Examples of such explanations by intuitionists as Gödel is thinking of can be found in L.E.J. Brouwer, *Collected works. I: Philosophy and Foundations of Mathematics*, ed. A. Heyting (Amsterdam: North-Holland, 1975), p. 423 and p. 511.

> a satisfactory foundation of Cantor's set theory in its whole original extent and meaning, namely the axiomatics of set theory interpreted in the way sketched below.[146]

It is possible to have qualms with Gödel's characterization of intuitionism here. A potential unclarity in the use of "completely given" here is whether this may or may not involve certain (carefully controlled) idealizations: intuitionists do make such idealizations, and they are not finitists (as Gödel himself of course pointed out on other occasions).[147] And in any case, intuitionists accept essentially incomplete objects (choice sequences) in their ontology, which by their nature can never be completely given in either actual or (to intuitionistic standards) appropriately idealized intuition. For the contrast that Gödel wishes to draw here, these qualms do of course not make much difference. But the intuitionist can in any case point out the symmetry in the situation as sketched by Gödel in this passage, and use Gödel's own words to comment that the latter's realist views are "by no means a necessary outcome of a closer examination of [the] foundations [of Cantor's set theory and classical mathematics], but only the result of a certain philosophical conception of the nature of mathematics."

Thus, further argumentation is required. Immediately after presenting the argument from the psychological fact, Gödel introduces a second and apparently stronger argument:

> What, however, perhaps more than anything else, justifies the acceptance of this criterion of truth in set theory is the fact that continued appeals to mathematical intuition are necessary not only for obtaining unambiguous answers to the questions of transfinite set theory, but also for the solution of the problems of finitary number theory (of the type of Goldbach's conjecture), where the meaningfulness and unambiguity of the concepts entering into them can hardly be doubted. This follows from the fact that for every axiomatic system there are infinitely many undecidable propositions of this type.[148]

In particular, those undecidable propositions can take the simple form of Diophantine equations. It is through mathematical intuition that we come to see that these propositions are actually true (provided we believe in the consistency of the axioms). The intuitionist agrees, but would balk at taking the range of this intuition to be so large as to include transfinite set theory. According to the intuitionist, even if such an axiom can be shown to be consistent, there is no reason to assume that it is true.

In a third type of argument, not presented in the Cantor paper but as pertinent to the issue discussed there, Gödel aims to show not so much that intuition exists at all, but rather that, if one accepts its existence, it would be wrong to think of it as an all-or-nothing phenomenon that, for that very reason, would cancel any purported intuitive evidence for objects that is not maximal (as in most people's experience

---

[146] *CW III*, pp. 257–258.

[147] On idealizations in intuitionism, see M. van Atten, "Intuitionistic Remarks on Husserl's Analysis of Finite Number in the Philosophy of Arithmetic," *Graduate Faculty Philosophy Journal*, 25(2) (2004), pp. 205–225.

[148] *CW II*, pp. 268–9.

the evidence for transfinite sets is not). Rather, the evidence that intuition provides comes in degrees:

> We have no absolute knowledge of anything. There are degrees of evidence. The clearness with which we perceive something is overestimated. The simpler things are, the more they are used, the more evident they become. What is evident need not be true. If 10^{10} is already inconsistent, then there is no theoretical science.[149]

It should not be surprising that what Gödel says here—to Wang, in the 1970s—is in agreement with Husserl's views on intuition, views that Gödel began to study in 1959. In an unpublished manuscript titled "For Perspicuous Objectivity," Hao Wang reports that in November 1972, Gödel drew his attention to the following sentence in Husserl's essay "Philosophy as rigorous science": "Obviously essences can also be vaguely represented, let us say represented in symbols and falsely posited; then they are merely conjectural essences, involving contradiction, as is shown by the transition to an intuition of their inconsistency."[150] Gödel then commented that he was glad that Husserl also recognizes the possibility of error'.[151] A related way in which an essence can be represented vaguely (one which is equally important to the present discussion) is if this representation, is in part symbolic and in part intuitive. We then have evidence (in Husserl's sense) up to a certain degree; but the possibility of error remains.[152] On the other hand, as Husserl writes immediately after the sentence just quoted, '[i]t is possible, however, that their vague position will be shown to be valid by a return to the intuition of the essence in its givenness'.[153]

It is also in the recognition that intuition is not an either-or affair but comes in degrees (i.e., intuitions are, generally speaking, partial) that Gödel sees an answer to skepticism:

> We have no absolute knowledge of anything. To acknowledge what is correct in skepticism serves to take the sting out of skeptical objections. None of us is infallible. Before the paradoxes Dedekind would have said that sets are just as clear as integers.[154]

This remark about Dedekind points to a difference between the degree of evidence we have for the integers and that for sets. This can be generalized from different types of objects to different conceptions of mathematics as a whole and by

---

[149] Wang, *Logical Journey*, p. 302, 9.2.35.
[150] The edition that Gödel read with Wang is in E. Husserl, *Phenomenology and the Crisis of Philosophy*, trl. Q. Lauer (New York: Harper & Row, 1965), p. 112.
[151] Wang, *For Perspicuous Objectivity. Discussion with Gödel and Wittgenstein*, (unpublished manuscript), p. 67.
[152] An extensive account by Husserl on how knowledge of an object is in general founded on both intuitive and non-intuitive presentations of that object is in his sixth *Logische Untersuchung* in E. Husserl, *Logische Untersuchungen. Zweiter Band, 2. Teil* Husserliana vol. XIX/2 (Den Haag: Martinus Nijhoff, 1984). See also Husserl's considerations on adequacy and apodicticity of knowledge at the beginning of his *Cartesianische Meditationen* (a book Gödel valued). E. Husserl, Cartesianische Meditationen und Pariser Vorträge Husserliana vol. I (Den Haag: Martinus Nijhoff, 1950).
[153] E. Husserl, *Phenomenology and the Crisis of Philosophy*, p. 113.
[154] Wang, *Logical Journey*, p. 302, 9.2.35.

doing so, one may obtain a scale such as the following one suggested by Gödel in a version of his paper on Carnap paper:

> The field of unconditional mathematical truth is delimited very differently by different mathematicians. At least eight standpoints can be distinguished. They may be characterized by the following catchwords: 1. Classical mathematics in the broad sense (i.e., set theory included), 2. Classical mathematics in the strict sense, 3. Semi-Intuitionism, 4. Intuitionism, 5. Constructivism, 6. Finitism, 7. Restricted Finitism, 8. Implicationism.[155]

That Gödel interpreted such a list as a scale of evidence is clear from the following:

> Without idealizations nothing remains: there would be no mathematics at all, except the part about small numbers. It is arbitrary to stop anywhere along the path of more and more idealizations. We move from intuitionistic to classical mathematics and then to set theory, with decreasing certainty. The increasing degree of uncertainty begins [at the region] between classical mathematics and set theory. Only as mathematics is developed more and more, the overall certainty goes up. The relative degrees remain the same.[156]

Moreover, Gödel claims that it would be arbitrary to draw (as the intuitionist does) a line that partitions the scale into an acceptable and an unacceptable part:

> Strictly speaking we only have clear propositions about physically given sets and then only about simple examples of them. If you give up idealization, then mathematics disappears. Consequently it is a subjective matter where you want to stop on the ladder of idealization.[157]

Behind accepting this role of subjectiveness must be the idea that the various possible idealizations are in a sense continuous with one another; and, for Gödel, even as evidence decreases with each further idealization, there is still intuition of a sufficiently high degree which remains to give a purchase on the transfinite. An intuitionist, however, will argue that it is not at all arbitrary where to stop idealizing, and that the place to stop occurs well before having reached all of classical mathematics in the strict sense (let alone transfinite set theory). An argument to this effect is found in the inaugural address of Brouwer's pupil Arend Heyting from 1949. He noticed that within the intuitionistic school, there is disagreement as to what idealizations are permitted. The disputed notions are those of negation, choice sequences, and certain proof methods (Brouwer accepted all of these). Each depends on certain idealizations; with every new idealization 'we descend a step on the stairs of clarity', as Heyting says. He then wonders if, once you are prepared to descend those stairs at all, this might not provide a justification of classical mathematics:[158]

---

155 *CW III*, p. 346 n32.
156 Wang, *Logical Journey*, p. 217, 7.1.11.
157 Wang, *Logical Journey*, p. 217, 7.1.10.
158 Also compare Heyting's passage to Gödel's lecture from 1933, "The present situation in the foundations of mathematics," *CW III*, **1933o*, p. 51. Gödel presented it to a meeting of the Mathematical Association of America in Cambridge, MA, on December 30, 1933 (which meeting Heyting did not attend).

> Those concepts and methods that are not accepted by all intuitionistically oriented mathematicians, upon introspection turn out to have different degrees of clarity, and to be accompanied by convictions of correctness of different intensities. It lies close to hand to ask whether it isn't then also the case that much can be said for accepting the independent existence of the mathematical objects, independent of our thinking, and thereby arrive at classical mathematics, even if this diminishes the clarity of the concepts a bit. For me, the answer to this question is that this step is not at all comparable to the earlier ones. So far, we remained in the realm of mental constructions, now we would all of a sudden leave that. We would be asked not merely to accept a new construction method, but a philosophical thesis, of which the sense and motivation are questionable already for the objects in daily life and which would become even less understandable when applied to mathematical objects. The stairs that slowly led down from the light of day into the darkness stops here, and the next step would be a jump into the darkness of an bottomless well.[159]

This passage illustrates Gödel's general point that "There is a choice of how much clarity and certainty you want in deciding which part of classical mathematics is regarded as satisfactory: this choice is connected to one's general philosophy.";[160] but Heyting warns that it is not just a question of "how much" and that there are also qualitative differences between the idealizations that introduce discontinuities in the scale. These discontinuities put one's notion of intuition under pressure (and hence the range of idealizations one can reasonably make on the basis of that notion).

We will not discuss possible intuitionistic criticisms of Gödel's three arguments any further, but note that the task of their assessment emphasizes the need for further study of the notion of intuition—which notion, for all their differences, is central to both Brouwer's intuitionism and Gödel's platonism—before it can be made to do the work Gödel wants it to do towards, e.g., a decision of the continuum hypothesis. Phenomenology is a systematic way of going about such a study;[161] and indeed, an unpublished draft of the 1964 supplement to the Cantor paper contains an additional paragraph at the end that starts, "Perhaps a further development of phenomenology will, some day, make it possible to decide questions regarding the soundness of primitive terms and their axioms in a completely convincing manner."[162] Of course, Gödel's intention when writing this will not primarily have been to settle the dispute with the intuitionist; but a development as Gödel hopes for here may well have that effect.

5.2.2 Success

> Secondly, even disregarding the intrinsic necessity of some new axiom, and even in case it has no intrinsic necessity at all, a probable decision about its truth is possible also in another way, namely, inductively by studying its "success." Success here means fruitfulness in consequences, in particular in "verifiable" consequences, i.e. consequences verifiable without

[159] A. Heyting, *Spanningen in de wiskunde* inaugural lecture, University of Amsterdam, 1940 (Groningen: Noordhoff, 1949), p. 13; translation ours.
[160] Wang, *Logical Journey*, p. 216, 7.1.9.
[161] *CW III*, **1961/?*, p. 383.
[162] GN, item 040311, p. 12.

> the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover, and make it possible to contract into one proof many different proofs. The axioms for the system of real numbers, rejected by the intuitionists, have in this sense been verified to some extent, owing to the fact that analytic number theory frequently allows one to prove number-theoretical theorems which, in a more cumbersome way, can subsequently be verified by elementary methods. A much higher degree of verification than that, however, is conceivable. There might exists axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding such powerful methods for solving problems, (and even solving them constructively, as far as that is possible) 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.[163]

Because of its dependence on (cognitive, aesthetic) values and choices of criteria, the notion of success is much less tractable than that of intuition, which stands in a direct relation to its object. On the other hand, indications of success may be easier to recognize than sharp intuitions, which will generally take much care and effort to arrive at. Some of the precautions that should be taken when searching for intuitions of essences are described by Husserl in sections 80–98 of *Erfahrung und Urteil*.[164] Gödel observed that "appealing to intuition calls for more caution and more experience than the use of proofs—not less";[165] and, as Wang says, "this may be the reason why one could believe in this strong position [according to which we perceive mathematical objects that exist objectively] and yet not regard the criterion of pragmatic success as entirely superfluous."[166]

A clear case of a conflict of values is presented by $V{=}L$. If one defines success of an axiom in terms of its power to decide questions, then it seems $V{=}L$ should qualify for that reason. But the axiom conflicts with certain large cardinal axioms that are believed to be consistent. These may not (yet) be seen to be themselves intrinsically necessary, but they too are successful in Gödel's sense, in that they decide many questions lower down (Diophantine equations).[167] Such a conflict of values can be resolved by an appeal to intuition, for example if intuition of the essence of set yields that all consistent sets should exist;[168] then, if one believes the large cardinal axioms in question to be consistent, these cardinals therefore exist. A related argument is that $V{=}L$ may be successful but is also conceptually impure (a judgement based on intuition), in the sense explained in Section 3 above.

There clearly is an asymmetry between the two criteria of intuition and success. As a first approximation one could say that intuition is conclusive while induction is not; but this approximation needs refinement once we take the following into

[163] *CW II*, p. 261.
[164] E. Husserl, Erfahrung und Urteil (Hamburg: Meiner 1985).
[165] Wang, *Logical Journey*, p. 301.
[166] Wang, *From Mathematics to Philosophy*, p. 207.
[167] *CW II*, p. 261. Note that the large cardinal axioms do not decide "lower down" statements like the *CH*, as, roughly speaking, large cardinal properties are preserved under Cohen extensions, which of course change the value of the continuum. See D.A. Martin, "Hilbert's First Problem: The Continuum Hypothesis," *Proceedings of the Symposia in Pure Mathematics*, vol. 28 (1976).
[168] See the sections on maximality and generic absoluteness, below.

account. As we saw above, Husserl emphasizes that in the correct understanding of intuition, it is not an either-or phenomenon – evidence comes in degrees. On such a construal, specific intuitions will in general not be conclusive in an apodictic sense. This it has in common with inductive arguments, but there are essential differences. Intuitions are of the objects (and the states of affairs composed of them) directly. By their (horizonal) structure, intuitions lend themselves to explication; they suggest ways in which they can be unfolded.[169] Inductive arguments (in terms of fruitful consequences) on the other hand are indirect; evidence for them will not come from the objects (or states of affairs) themselves but from seeing the truth of their consequences. This induces a principled epistemological difference because inductive arguments may have heuristic value but cannot have the same regulative significance (towards the ideal of full clarity and insight, through explication) as partial intuitions. Moreover, this emphasizes the foundational priority of intuition as such over induction as such: for as Wang remarks, "The truth of these consequences, however, had also be seen by mathematical intuition, and we see certain mathematical propositions, such as numerical computations, to be true directly, without going through the axioms. Indeed, we apply our intuition at all levels of generality."[170]

An example of an axiom that extends ZFC and that can be defended on the basis of its success is the axiom stating the existence of inaccessibles:

> A closely related fact is that the assertion (but not the negation) of the axiom [of inaccessibles] implies new theorems about integers (the individual instances of which can be verified by computation). So the criterion of truth explained on page 264 is satisfied, to some extent, for the assertion, but not for the negation. Briefly speaking, only the assertion yields a "fruitful" extension, while the negation is sterile outside of its own very limited domain.[171]

This is a reference to the fact that in the theory ZFC + "there is an inaccessible cardinal" one can prove the statement "there exists a model of ZFC." But this is equivalent to Con(ZFC), by the completeness theorem, which is a Π^0_1 number-theoretic or Diophantine statement. As we saw above, Gödel also had a (much more conclusive) justification of inaccessibles from an analysis of the concept of set. Thus inaccessibles are justified in two very different ways; Gödel probably wanted to include the pragmatic argument because it may serve to convince those who are more easily impressed by formal(istic) considerations.[172]

In the early 1970s Gödel suggested to Wang that the existence of measurables may be perhaps be verified in the same way:

[169] E.g. Husserl, *Ideen I*, section 44; *Cartesianische Meditationen*, Section 19; *Erfahrung und Urteil*.

[170] Wang, *Logical Journey*, p. 244.

[171] *CW II*, p. 267.

[172] Note that Gödel's page reference is to the original publication, whose page numbers are indicated in the margins of the Collected Works. He introduces the argument from success which we quoted at the beginning of this section only on page 265. Yet there is clearly a pragmatic aspect to this second argument for inaccessibles. In Wang's chapter "The Concept of Set," which was written in close collaboration with Gödel, this is made explicit (*From Mathematics to Philosophy*, p. 203).

> The hypothesis of measurable cardinals may imply more interesting (positive in some yet to be analyzed sense) universal number-theoretical statements beyond propositions such as the ordinary consistency statements: for instance, the equality of p_n (the function whose value at n is the n-th prime number) with some easily computable function. Such consequences can be rendered probable by verifying large numbers of numerical instances.[173]

On the other hand, *CH* itself cannot have new number-theoretic consequences, as is clear from absoluteness of arithmetical statements: arithmetical statements are true in a given model M of ZFC if and only if they are true in L over M. (Kreisel[174] draws attention to the fact that Gödel did not realize this in Section 3 of the 1947 paper: "Certainly [...] mere consistency leaves open the possibility that CH has new, even false arithmetic consequences; but a glance at his own definition of L [...] shows that *CH*, and even $V = L$ has none at all. Gödel's oversight is natural enough if consistency is regarded as an end in itself.")

It turned out (at about the time Gödel was writing the 1964 revision and the 1966 postscript to it) that the negation of CH does not have new number-theoretic consequences either, as a simple observation about forcing shows. Hence, different aspects of fruitfulness than arithmetic consequences (verifiable up to any given integer) will have to be taken into account if an argument from success is to lead to a probable decision of *CH*.

Gödel mentions several different aspects of fruitfulness. One is that of making proofs simpler and "easier to discover," i.e., more obvious; where without the new axiom their proofs are long and "cumbersome." It is an open question whether, to the extent that fruitfulness of this kind is indeed truth-conducive, what is operative here is a relation between truth and aesthetic factors, or between truth and exigence of resources, or both (on the possible ground that economy of resources may be, at least in part, itself an aesthetic factor). So far, no convincing arguments of this type are known that suggest a decision of *CH*. Certainly the recently proposed solution to *CH* we will discuss below is by no means "easy," "obvious," or in some other sense elementary. Should the proof eventually work, then it is of course to be expected that in due course simplifications (e.g., from conceptual insight) will be found.

In the supplement to the 1964 edition of the Cantor paper, Gödel points out that his second truth criterion has not (at the time of writing) led to concrete results:

> It was pointed out earlier [...] that, besides mathematical intuition, there exists another (though only probable) criterion of the truth of mathematical axioms, namely their fruitfulness in mathematics and, one may add, possibly also in physics. This criterion, however, though it may become decisive in the future, cannot yet be applied to the specifically set theoretical axioms (such as those referring to great cardinal numbers), because very little is known about their consequences in other fields. The simplest case of an application of the criterion under discussion arises when some set-theoretical axiom has number theoretical consequences verifiable by computation up to any given integer. On the basis of what is known today, however, it is not possible to make the truth of any set-theoretical axiom reasonably probable in this manner.[175]

---

[173] Wang, *Logical Journey*, p. 263, 8.3.13.
[174] Kreisel, "Kurt Gödel," p. 197.
[175] *CW II*, p. 269.

It is not entirely clear what Gödel has in mind when he first says that the second criterion is "only probable" and then goes on to add that this inductive criterion "may become decisive in the future." We already quoted from an unpublished draft of this supplement, in which there is an additional final paragraph. In full, it reads:

> Perhaps a further development of phenomenology will, some day, make it possible to decide questions regarding the soundness of primitive terms and their axioms in a completely convincing manner. As of now it seems to me that the character of cogency of its axioms[176] and the success of its development are sufficient reasons for putting trust in Cantor's set theory, i.e., in mathematics in its whole extension.[177]

Gödel expects phenomenology to lead to decisions; in contrast, cogency of the axioms (clearly meant in a sense that does not amount to complete clarity obtained by phenomenological analysis) and inductive arguments from success do not go that far, but increase confidence. Of course, the phenomenological investigation into the basic concepts of mathematics will not by itself settle mathematical questions. As Gödel said to Wang in the 1970s,

> The epistemological problem is to set the primitive concepts of our thinking right. For example, even if the concept of set becomes clear, even after satisfactory axioms of infinity are found, there would remain technical (i.e. mathematical) questions of deciding the continuum hypothesis from the axioms.[178]

Some years after Gödel had written about this notion of success and the senses in which axioms can to some extent be verified, specific examples were found. We mention the following two.

A result obtained by Richard Laver in 1992 says that, if one assumes the existence of very large cardinals (larger than supercompact), one can find a decision method for the word problem of the free algebra with one generator and one left-distributive binary operator.[179] Later, Patrick Dehornoy gave a proof without the large cardinals.[180] So the large cardinals gave the correct result, which in a sense verified their use.

---

[176] footnote Gödel: "On the basis of the concept of set explained on p. ," where he will have meant to refer to his explanation of the iterative conception; in the Collected Works, this is on pp. 258–259.

[177] GN, item 040311, p. 12.

[178] GN: item 013183.5, dated March 11, 1976: a list of quotations from Gödel, corrected by himself.

[179] R. Laver, "The left distributive law and the freeness of an algebra of elementary embeddings," *Advances in Mathematics* 91(2) (1992), pp. 209–231.

[180] P. Dehornoy, "Elementary embeddings and algebra." Invited chapter in the forthcoming *Handbook of Set Theory*, eds. M. Foreman, A. Kanamori, and M. Magidor.

Another, dramatic example of this has to do with Borel Determinacy, which was proved to follow from the ZFC axioms only by D.A. Martin in 1975,[181] where an earlier proof also due to Martin[182] used a measurable cardinal.[183]

In this way Borel Determinacy is a "verifiable consequence," in Gödel's sense of the phrase here, i.e., it was proved without using measurables, and the measurables in turn were verified by their having lead to the "correct" result.[184] As Yiannis Moschavakis put it in his book *Descriptive Set Theory*: "This important result of Martin answered a long-standing question and *lent considerable respectability to the practice of adopting determinacy hypotheses*" [emphasis ours]. [185]

5.3 Gödel's Interpretation of Cohen's Independence Result

In 1963, Paul Cohen proved the consistency of $\neg CH$ with ZFC and thereby, as Gödel had shown the consistency of *CH* with ZFC, the independence of *CH* from ZFC. Cohen sent his proof to Gödel, who (in a draft letter which may or may not have been sent) commented, "Reading your proof had a similarly pleasant effect on me as seeing a really good play."[186]

As we had occasion to mention elsewhere,[187] to Gödel this independence of *CH* from ZFC did not, by suggesting a certain relativism in set theory, pose a threat to his realism. To Church, who did interpret the independence proof along these lines, he wrote on September 29, 1966:

> You know that I disagree [i.e. with Church] about the philosophical consequences of Cohen's result. In particular I don't think realists need expect any permanent ramifications (see bottom of p. 8) as long as they are guided, in the choice of the axioms, by mathematical intuition and by other criteria of rationality.[188]

The fact that (like many others) Church was so impressed by the independence of *CH* from ZFC as to conclude from it to a kind of relativism in set theory may well

---

[181] D.A. Martin, "Borel Determinacy," *Annals of Mathematics* 102 (1975), pp. 363-371.
[182] D.A. Martin, "Measurable Cardinals and Analytic Games," *Fundamenta Mathematicae* 66 (1969/1970), pp. 287–291.
[183] We note also here Jeff Paris's result that Pi^0_4-determinacy is provable in ZFC. Paris, J. B. "ZF $\vdash \sum^0_4$ determinateness," *The Journal of Symbolic Logic* 37 (1972), pp. 661–667.
[184] The further, aesthetic issue, having to do with simplicity, we do not comment on here. The axiom asserting the existence of a measurable was simply fruitful in yielding a powerful result about Borel games.
[185] Y. Moschavakis, *Descriptive Set Theory* (Amsterdam: North-Holland 1980), p. 357.
[186] *CW IV*, p. 378.
[187] Van Atten and Kennedy, "On the Philosophical Development of Kurt Gödel," p. 470.
[188] *CW IV*, p. 372. Gödel's page reference is to the text of Church's talk "Paul J. Cohen and the Continuum Problem," published in 1968 in the Proceedings of the International Congress of Mathematicians (Moscow-1966), pp. 15–20; p. 8 of the manuscript corresponds to p. 18 of the publication.

have had its ground in the unprecedented scope and success of ZFC as a foundation of classical mathematics.

For Gödel, however, there is no fundamental relativity in set theory. Set theory describes a certain (part of abstract) reality, which is therefore the theory's intended model. Cognitive access to that reality is provided by mathematical intuition. Like Brouwer, Gödel holds that mathematical intuition is separate from language, and that language has a practical but not fundamental role to play in obtaining intuitions of mathematical objects:

> Language is useful and even necessary for fixing our ideas. But this is a purely practical affair. Our mind is more inclined to sensual objects, which help to fix our attention on abstract objects. This is the only importance of language.[189]

Proofs that certain statements (or classes of statements) are independent of a given formalism, or that given a given formalism admits of non-intended models, or that formalisms as such have intrinsic limitations, do not, in principle, stand in the way of our mind's capacity to obtain knowledge of the mathematical realm. This explains Gödel's remark to Church about Cohen's proof. To Gödel, the independence of *CH* from (for example) ZFC is no reason whatsoever to doubt that a decision of *CH* is in principle possible. These considerations are not meant to suggest that Gödel did not credit the practical value of formal systems and formalization. In what is now referred to as his "lecture at Zilsel's," held in 1938, he acknowledged that "If the original Hilbert program could have been carried out, that would have been without any doubt of enormous epistemological value."[190] (Note that he does not add "ontological"; it is not immediate that for a realist the epistemological success of Hilbert's program would also have entailed an ontological reduction.[191]) That program of course did not succeed in its original form, but there were other benefits to formal systems; in a letter of December 1918, 1968, to Dana Scott, Gödel remarks:

> In my opinion the formalistic spirit is *extremely important* for mathematics as a technique of solving problems. Also, I perfectly agree with you that formalization, in practice, is an *indispensable* aid to understanding.[192]

5.4 Maximality

As we have seen above, Cohen's result cleared a path for considering axioms deciding CH along grounds which were no longer constrained by knowing only that

---

[189] Wang, *Logical Journey*, p. 180.

[190] *CW III*, p. 113.

[191] This may explain why Gödel could, as we saw, in 1938 say that a successful completion of Hilbert's program would have been "of enormous epistemological value," and in (or around) 1947 decline an invitation to write on Hilbert's program by saying that he was not sufficiently sympathetic to it (*CW II*, p. 144).

[192] GN, item 012279; original emphasis.

CH was consistent. From a practical point of view this means that axioms implying $\neg CH$ are going to be "in play" in a way they would not have been otherwise.

Though as we have seen, Gödel was interested in axioms of many different kinds, as far as his concrete attempts to solve the continuum problem however, in what can be said to be his last work of a technical nature, it was to Hausdorff's scale axioms that he turned (see the next section). These are prima facie maximality principles, in that some maximal family of functions is asserted to exist, relative to the particular scale involved. But they are at the same time of a somewhat different flavor from what were thought to be maximality principles at the time (see below).

The "set of" operation on which the iterative concept of set is based is "opened forward," so to speak. In footnote 23 of the 1964 version of the paper, while discussing the idea that *CH* might be decided on the basis of axioms about definability p. 262, Gödel remarks that this is known for the type of definability known as constructibility:

> On the other hand, from an axiom in a sense opposite to this one, the negation of Cantor's conjecture could perhaps be derived. I am thinking of an axiom which (similar to Hilbert's completeness axiom in geometry) would state some maximum property of the system of all sets, whereas axiom A [$V=L$] states a minimum property. Note that only a maximum property would seem to harmonize with the concept of set explained in footnote 14.[193]

Footnote 14, as we saw above, asserts that the existence of multitudes is independent of whether we can define them in a finite number of words or whether they are random. So sets are what they are independent of their definitions (which goes some way toward arguing against restricting the notion of set to that of constructible set). Gödel is explicit that a maximality property of this type "harmonizes" with the basic concept. It is not clear that the choice of that particular word should be taken to mean that such a property is not actually contained in the pure concept of set. Recall that although $V=L$ (which is quite the opposite of a maximality principle) is in a way a natural extension of ZFC, its particular naturalness may not persist in a further unfolding of the concept of set.

This period saw Gödel begin to concentrate on maximality principles. Already in the late 1950s Gödel wrote to Stanislaw Ulam about a maximality principle of von Neumann:

> The great interest which this axiom has lies in the fact that it is a maximality principle, somewhat similar to Hilbert's axiom of completeness in geometry. For, roughly speaking, it says that any set which does not, in a certain well defined way, imply an inconsistency exists. Its being a maximum principle also explains the fact that this axiom implies the axiom of choice. I believe that the basic problems of set theory, such as Cantor's continuum problem, will be solved satisfactorily only with the help of stronger axioms of *this* kind, which in a sense are opposite or complimentary to the constructivistic interpretation of mathematics.[194]

---

193 *CW II*, pp. 262–263.

194 S. Ulam, "John von Neumann, 1903–1957," *Bulletin of the American Mathematical Society* 64(3) (1958), part 2, pp. 1–49, as quoted in *CW II*, p. 168; original emphasis. Note that this is different from the very similar passage in *CW V*:295.

To Wang, Gödel reaffirmed this assent to von Neumann's principle by saying that for him, completeness means that every consistent set exists.[195]

6 1970–1975: Gödel's Concrete Attempts to Settle *CH*

In the 1970s Gödel made several efforts at deciding *CH*. These were based on axioms (and axiom schemas) about infinite sequences of integers and scales of functions. We will not go into the technical details here, but give a brief account of this episode.[196]

From a letter he wrote to Cohen in January 1964,[197] it becomes clear that Gödel had the basic idea for one of these axioms already then. "I always suspected that, in contrast to the continuum hypothesis, this proposition is correct and perhaps even demonstrable from the axioms of set theory." Whether the fact that a few years later he took this to be an axiom reflects a failed attempt to prove it or simply a shortcut we do not know. Neither do we know whether the scale axioms were the outcome of a conceptual analysis of the kind described in 1947 and 1964.

Gödel wrote a manuscript based on the scale axioms titled "Some considerations leading to the probable conclusion that the true power of the continuum is $\aleph_2$." He planned to submit it to the *Proceedings of the National Academy of Sciences* and in 1970 first sent it to Tarski to solicit comments. D.A. Martin found that the argument was incorrect, as it contradicted a result of Robert Solovay. Tarski sent the paper back to Gödel, simply saying that he would soon hear more about it.

Gödel then wrote another manuscript, this time not meant for publication but for his own use ('nur für mich geschieben'), titled "A proof of Cantor's continuum hypothesis from a highly plausible axiom about orders of growth." In it, he said that the argument presented there "gives *much* more likelihood to the truth of Cantor's continuum hypothesis than any counterargument set up to now gave to its falsehood."[198] Remarkably, and for the first time, Gödel had at least briefly convinced himself that *CH* is true instead of false. The manuscript establishes, correctly, an equivalence between a particular instantiation of one of the axiom schemas and *CH*.

In an unsent letter to Tarski of 1970, Gödel said about the paper he had sent him that

> Unfortunately my paper, as it stands, is no good. I wrote it in a hurry shortly after I had been ill, had been sleeping very poorly and had been taking drugs impairing the mental functions. [...] My conviction that $2^{\aleph_0} = \aleph_2$ of course has been somewhat shaken. But it still seems plausible to me. One of my reasons is that I don't believe in any kind of irrationality such as, e.g., random sequences in an absolute sense.[199]

---

[195] Wang, *Logical Journey*, p. 144.
[196] For the details, see *CW II*, pp. 173–175 and *CW III*, pp. 405–420.
[197] *CW IV*, pp. 383–384
[198] *CW III*, p. 423.
[199] *CW III*, p. 424.

Interestingly, $\aleph_2$ is also the value that some set theorists today think is correct (see below).

According to the diaries of Gödel's friend Oskar Morgenstern, Gödel kept working on the problem, apparently again to show that the power of the continuum is $\aleph_2$; in what is presumably the last of Morgenstern's reports, from late 1975, Gödel had become convinced of having a correct proof, according to which the power of the continuum was "*not* $\aleph_2$, but rather 'different from $\aleph_1$.'"[200]

From Gödel's first manuscript a correct proof for a weaker proposition has been reconstructed by Paul Larson, Jörg Brendle and Stevo Todorcevic.[201] They have formulated three new axioms that are implicit in Gödel's original paper that together imply that $2^{\aleph_0} \leqslant \aleph_2$. The problem is that there is no reason to find these new axioms self-evident and accordingly they have not led to a new credible extension of ZFC.

7 A Theme in Contemporary Set Theory

It is a very striking fact that today one begins to encounter the view, by no means unanimously held but nevertheless expressed by a number of set theorists, that the *CH* is close to being solved, or if not, is once again considered open. Already in 1980, Kreisel noted that the suggestion that *CH* may fail to have a definite truth value for the intended interpretation at all—a suggestion motivated by lack of progress in spite of many attempts from different directions—simply overlooks that there are "infinitely many false starts, perhaps due to a systematic oversight, for any problem."[202] In 2000 Woodin echoed Kreisel's observation:

> There is a tendency to claim that the Continuum Hypothesis is inherently vague and that this is simply the end of the story. But any legitimate claim that *CH* is inherently vague must have a mathematical basis, at the very least a theorem or collection of theorems. My own view is that the independence of *CH* from ZFC, and from ZFC together with large cardinal axioms, does not provide this basis [...] Instead, for me, the independence results for *CH* simply show that *CH* is a difficult problem.[203]

7.1 Generic Absoluteness

The main method for studying the incompleteness phenomenon in set theory is (at present) the forcing method. Forcing does not change the first order arithmetic of integers; the arithmetic of integers is "forcing absolute," i.e., any arithmetic statement

200 For this paragraph, see Dawson, *Logical Dilemmas*, pp. 235–236.
201 J. Brendle, P. Larson, and S. Todorcevic, "Rectangular Axioms, Perfect Set Properties and Decomposition." Preprint at http://www.math.toronto.edu/larson/goedel.pdf
202 Kreisel, "Kurt Gödel," p. 212.
203 W.H. Woodin, "The Continuum Hypothesis" to appear in the *Proceedings of the Logic Colloquium 2000*, Paris.

true in a model (transitive) M of ZFC remains true in any generic extension $M[G]$ of M. Strictly speaking this is merely a consequence of the fact that forcing does not introduce new ordinals. As a consequence, no arithmetic statement can decide or be decided by the *CH*. For, suppose some arithmetic sentence ϕ did decide the *CH*. Then if we extend a model of *CH* to one in which *CH* is false, by forcing, then ϕ should no longer hold in the forcing extension, contradicting the forcing absoluteness of ϕ. A similar argument shows that the converse also holds, i.e. neither the *CH* nor its negation has arithmetic consequences.

This is contrary to what Gödel conjectured in the 1964 supplement to Cantor paper, namely:

> The generalized continuum hypothesis, too, can be shown to be sterile for number theory ... whereas for some *other* assumption about the power of $2^{\aleph_\alpha}$ this is perhaps not so.[204]

However Gödel notes the forcing argument in the postscript to that paper:

> Shortly after the completion of the manuscript of the second edition of this paper the question whether Cantor's continuum hypothesis is decidable from the von Neumann-Bernays axioms of set theory (the axiom of choice included) was settled in the negative by Paul J. Cohen ... It turns out that for all $\aleph_\tau$ defined by the usual devices and not excluded by König's theorem ... the equality $2^{\aleph_0} = \aleph_\tau$ is consistent in the weak sense *(i.e. it implies no new number-theoretical theorem).*[205]

The forcing absoluteness of arithmetic statements is perhaps an explanation of what may be called the "empirical completeness" of ZFC as far as arithmetic is concerned; the idea behind this being that incompleteness phenomena are reduced to "residual incompleteness" or incompleteness arising from ad hoc sentences such as the Gödel sentences.[206] That no new, non ad hoc independent arithmetical statements have emerged is evidence for empirical completeness.[207]

CH of course is a different matter. Not only it is not a question about integers, it is even not about reals,[208] but about sets of reals. As forcing determines statements like *CH* in various ways, any attempt to find an extension of ZFC which settles *CH* will have to address the "essential variability in set theory"[209] due to forcing—it must attack forcing head-on, so to speak.

In forcing, the class of formulas taken and the type of forcing are parameters and can be varied. The contemporary notion of generic absoluteness studies the preservation of different classes of formulas under different forcings.

---

204 Italics the authors'.

205 Italics the authors'.

206 A different view has been expressed by Paul Cohen, who has stated the view that any interesting statement about primes will eventually be shown to be independent of ZFC.

207 But see Harvey Friedman's work on structurally simple combinatorial statements which are equivalent to large cardinals.

208 As a matter of interest in this connection it is now known that there are two models of set theory with the same ordinals, cardinals and reals, one satisfying *CH* and the other not. Folklore, communicated by Matt Foreman.

209 P. Dehornoy, "Recent Progress on the Continuum Hypothesis (After Woodin)." http://www.math.unicaen.fr/~dehornoy/surveys.html

The first formulation of a general generic absoluteness principle for other than arithmetic statements, is due to Jonathan Stavi and Jouko Väänänen and dates from the late 1970s.[210] Note that the idea that any set that can exist, does exist, may turn out to be inconsistent if the meaning of the highly ambiguous "can" is not carefully specified; for example, both the *CH* and its negation state the existence of sets which can exist in a generic extension. Stavi and Väänänen state the following principle: any formula with parameters of hereditary cardinality less than the continuum that can be made true by c.c.c. forcing and that cannot be falsified later by c.c.c. forcing, is already true. The approach was motivated by the study of generalized quantifiers and the idea that the continuum is or should be "as large as possible." One of the main results in that paper is that Martin's axiom is equivalent to a very natural weakening of this. This was discovered independently by Joan Bagaria.[211]

The maximality principle "what can be forced and remains true in further forcing, is true" was rediscovered in 2003 by Joel Hamkins[212] following an idea found in 1999 by C. Chalons, a doctoral student of Boban Velickovic in Paris, whose work has remained unpublished, apart from an electronic announcement.[213]

Finally, some fundamental results pertaining to generic absoluteness, due to Hugh Woodin, are cited below.

7.2 The Woodin Program

The approach associated with the Woodin school seeks to identify the largest possible class of forcing immune statements. In pursuit of this goal they have identified extensions of ZFC which decide a large family of statements, including *CH* itself.[214]

A key result was obtained in 1984, when Woodin proved,[215] based on work of Foreman, Magidor and Shelah,[216] that the first order theory of the structure $(H(\omega_1), \varepsilon)$ is invariant under any kind of forcing, relative to a certain large cardinal assumption, namely the existence of a proper class of so-called Woodin cardinals. Here $H(\omega_1)$ is the set of hereditarily countable sets, i.e. the set of sets which are

---

210 J. Stavi and J. Väänänen, "Reflection principles for the continuum." In *Logic and Algebra*, ed. Yi Zhang, Contemporary Mathematics, vol. 302 (AMS 2002), pp. 59–84.

211 J. Bagaria, "A Characterization of Martin's Axiom in Terms of Absoluteness," *The Journal of Symbolic Logic* 62(2) (1997), pp. 366–372.

212 J.D. Hamkins, "A Simple Maximality Principle," *The Journal of Symbolic Logic* 68(2) (2003), pp. 527–550.

213 C. Chalons, "An axiom schema," circulated email announcement (1999).

214 W.H. Woodin, "The Continuum Hypothesis (I and II)," *Notices of the American Mathematical Society* 48(6) (2001), pp. 567–576, and 48(7) (2001), pp. 681–690.

215 See W. Hugh Woodin, "The axiom of determinacy, forcing axioms and the non-stationary ideal," Walter de Gruyter and co., Berlin, 1999.

216 M. Foreman, M. Magidor, and S. Shelah "Martin's Maximum, Saturated Ideals, and Nonregular Ultrafilters", *Annals of Mathematics* 127(1) (1988), pp. 1-47.

countable, all elements are countable, all elements of elements are countable, etc. $H(\omega_1)$ can be in effect identified with the set of all real numbers. This effectively lifts the forcing absoluteness of arithmetic statements in the presence of ZFC to forcing absoluteness of the theory of the reals, more exactly of $H(\omega_1)$ in the presence of sufficiently large cardinals.

The next step was to formulate an axiom which would lift this result from $H(\omega_1)$ to $H(\omega_2)$, the set of sets of hereditary cardinality at most ω_1. Thus the new goal is to guarantee the forcing absoluteness of the first order theory of $(H(\omega_2), \varepsilon)$, in the presence of some new axioms. Note that both ω and ω_1 are elements of $H(\omega_2)$ and therefore the theory of $H(\omega_2)$ decides *CH*. Namely, *CH* states the existence of a bijection between ω_1 and all reals. Such a bijection would be an element of $H(\omega_2)$. Thus the existence can be stated as a first order property of $H(\omega_2)$. This observation leads to an important point: no large cardinal axiom can fix the theory of $H(\omega_2)$, as we can always change the value of $2^{\aleph_0}$ without affecting large cardinals. So something different was needed.

This next step was carried out by the following theorem, due to Woodin: The theory of $H(\omega_2)$ is forcing absolute relative to the theory ZFC + an axiom that Woodin calls the '$(\star)$-axiom'.[217] Moreover, the $(\star)$-axiom implies $2^{\aleph_0} = \aleph_2$. Whatever the $(\star)$-axiom is, Woodin proves that the mere existence of an axiom which fixes the theory of $H(\omega_2)$ in (Woodin's Ω-logic) violates the *CH*.

Finally, more recently Woodin has extended his approach to an argument against formalism and the view that the truth or falsity of *CH* has lost its meaning.[218]

It would be beyond the scope of this paper to indicate why Woodin's fundamental results, some of which we have cited, constitute a convincing solution to the continuum problem in the eyes of a number of set theorists. Clearly it is distinguished from other proposed extensions of ZFC, in that Woodin's extensions handle their forcing extensions already, whereas any other kind of canonical extension one may propose must still confront this kind of variability.

Interestingly, a number of different people have obtained results which point to the same conclusion about the value of the continuum. To cite just a few: Velickovic and Todorcevic have shown that the Proper Forcing Axiom PFA implies $2^{\aleph_0} = \aleph_2$.[219] PFA states that if P is proper, where the definition of properness is somewhat technical but generalizes both C.C.C. forcings and countably closed forcings, and D is a set of dense open subsets of P with $|D| \leqslant \aleph_1$, then there is a generic filter on P which meets every dense set in D. Justin Moore has recently shown that a very

---

217 W. Hugh Woodin, "The Continuum Hypothesis". I: *Notices of the American Mathematical Society* 48(6) (2001), pp. 567–576. II: *Notices of the American Mathematical Society* 48(7) (2001), pp. 681–690. Correction in 49(1) (2002), p.46.

218 See Coxeter Lectures, Fields Institute, Toronto, November 2002 at http://www.fields.utoronto.ca

219 B. Velickovic, "Forcing Axioms and Stationary Sets," *Advances in Mathematics* 94(2) (1992), pp. 256–284.

plausible bounded version of PFA, the so-called Bounded Proper Forcing Axiom BPFA implies the same.[220]

In the opinion of Kennedy, what is particularly interesting about BPFA in addition to its solving the continuum problem in the direction Gödel anticipated, is that it has the form of a reflection principle, principles which were very important to Gödel:

> All the principles for setting up the axioms of set theory should be reducible to Ackermann's principle: The Absolute is unknowable. The strength of this principle increases as we get stronger and stronger systems of set theory. The other principles are only heuristic principles. Hence, the central principle is the reflection principle, which presumably will be understood better as our experience increases. Meanwhile, it helps to separate out more specific principles which either give some additional information or are not yet seen clearly to be derivable from the reflection principle as we understand it now.[221]

In the opinion of Van Atten, on the other hand, reflection principles such as BPFA are not an example of what Gödel is speaking about here, for the following reasons: (1) Gödel here does not speak of reflection principles in the plural, but only about the most general principle 'The Absolute is unknowable'; (2) unlike BPFA, that principle is by its nature not completely formalizable (which makes possible the repeated application that Gödel speaks of here); (3) the general principle has, for a realist, an immediate plausibility on philosophical grounds (as was noted by Cantor, who first used it[222]) that cannot be claimed for BPFA.

It is in any case tempting to infer that the results such as the ones we have cited on PFA and BPFA are results Gödel would have ascribed clear significance to, including them under the category of inductive evidence. But a comprehensive review of recent results in set theory would be needed in order to evaluate whether these results are not anomalous.

There is a straightforward connection between a central aspect of Woodin's approach to *CH* and phenomenology. One of the ideas in Husserl's genetic analysis of judgement (as presented in *Formal and Transcendental Logic*, and in *Experience and Judgement*) is that the kinds of judgements that can legitimately be made in a domain of objects depend on the type of the objects. This leads to the idea that with different domains are associated different logics (which will be extensions of the minimal logic, defined by its applicability to all domains). (Robert Tragesser has developed this idea systematically in his book *Logic and Phenomenology*, mentioned above.) Woodin delineates the domain in which *CH* "lives," so to speak, and then searches for the most appropriate or most specific logic for that domain (more specific than first order logic which (classically speaking) is domain-independent). Thus, Woodin asks:

---

[220] "Set Mapping Reflection", *Journal of Mathematical Logic* 5(1) (2005), pp. 87–98; erratum in 5(2), p. 299.

[221] Wang, *A Logical Journey*, p. 283, 8.7.9.

[222] See note 2 in G. Cantor, 'Ueber unendliche, lineare Punktmannichfaltigkeiten. Nr.5', p. 205 in the reprint in his *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts* (E. Zermelo, ed.), Berlin: Springer, 1932.

> Can the theory of the structure $\langle H(\omega_2), \in\rangle$ be finitely axiomatized (over ZFC) in a (reasonable) logic which extends first order logic?[223]

One may even say that a positive answer to this question would be an instance of unfolding the concept of set, or the subconcept of set in the structure $H(\omega_2)$, in the sense that it determines what ways of reasoning (i.e. what logic) certain sets allow for. In other words, unfolding the notion of set need not result only in axioms (of the form "There exists a set with property P") but can also yield principles of (formal) reasoning.

The criticism of Woodin's project given so far[224] indicates that various supporting facts would need to be established in order for the solution to be universally, or perhaps, more, accepted. For example, the Ω-conjecture, a statement to the effect that Woodin's Ω-logic satisfies a natural completeness theorem, has not been proved.

Furthermore, any evaluation of the solution must take into account the privileged role that forcing plays in the construction of models of set theory. The Woodin program rests on the judgement that forcing is the only model construction technique to be considered; thus finding the canonical model for set theory means, principally, finding a reasonable way to "disable" forcing. Empirical completeness suggests that forcing plays this role already for the arithmetic statements. But this is an ambitious plan; in fact another one of Shelah's "Logical Dreams"[225], number 4.4, is: "Find additional methods for independence results (in addition to forcing and large cardinals/consistency) or prove the uniqueness of these methods." Empirical completeness suggests that this holds for the arithmetic statements. As no new independence results have come from other quarters sofar, it is reasonable to conjecture that forcing must be the only phenomenon which introduces variability.

We have suggested that Woodin's work is something of a beginning toward the project of reducing the variability in set theory to so-called residual incompleteness. But are there any convincing arguments for the notion that residual incompleteness really is "residual," that is, not a meaningful phenomenon mathematically?

We saw that Gödel gives an exact criterion for when the question of truth of an axiom A to be added to a theory T loses its meaning; just when $T + A$ and $T + \neg A$ are weak extensions of T, meaning definable in a ground model of T. It is interesting that Gödel does not consider the question whether one obtains strong or weak extensions from taking A to be $con(T)$. And perhaps if the terms weak and strong extension are interpreted loosely, $con(T)$ gives a weak extension, whereas $\neg con(T)$ does not.

---

[223] Woodin, "The Continuum Hypothesis. II," p. 682.

[224] See Matthew Foreman's "Has the continuum hypothesis been settled?", in Stoltenberg-Hansen, Viggo and Väänänen, Jouko (eds), *Logic colloquium '03. Proceedings of the annual European summer meeting of the Association for Symbolic Logic (ASL), Helsinki, Finland, August 14–20, 2003*. Wellesley, MA: A K Peters; Urbana, IL: Association for Symbolic Logic (ASL). Lecture Notes in Logic 24 (2006), pp. 56–75.

[225] S. Shelah, "Logical Dreams," *Bulletin of the American Mathematical Society* 40 (2003), pp. 203–228.

That is to say, adding solutions of new Diophantine equations, i.e., elements which witness statements about inconsistency, would give strong extensions. But this means there is an asymmetry present resembling the asymmetry induced when we considered extending the ZFC axioms by inaccessibles. So the question of whether an axiom of this type A is true, where A denotes the consistency of a theory T, is meaningful by Gödel's own criterion, and therefore perhaps not "residual." But this does not seem right. Believing in those particular (weak) extensions, i.e. assuming consistency, must be warranted in any case: consistency is the minimum assumption.[226] The insight that forms the basis of Feferman's analyses of hierarchies of reflection principles is also needed here: belief in the consistency of a set of axioms is supervenient on the belief in the axioms themselves. That is to say, believing in consistency should not commit us epistemologically to any principles beyond the axioms themselves.

This very brief sketch of Woodin's approach together with some of the criticism it has provoked is necessarily only too brief. The interested reader is referred to the literature.[227]

8 Other Developments

Once the independence of all non-trivial statements about the size of $2^{\aleph_0}$ were established by Cohen, attention immediately turned to other cardinals, e.g. to $2^{\aleph_n}$ and even to $2^{\aleph_\omega}$. It turned out that the power sets of regular cardinals like $\aleph_n$ could have any non-trivial cardinality and such statements were also independent from each other[228]. The case of singular cardinals remained a puzzle. For example, if $2^{\aleph_n} = \aleph_{n+1}$ for all n, does it follow that $2^{\aleph_\omega} = \aleph_{\omega+1}$? The Singular Cardinal Hypothesis, SCH, states that if κ is a singular strong limit cardinal then $2^\kappa = \kappa^+$. Ronald Jensen showed in a penetrating study[229] that SCH cannot be decided by forcing over L, or, more exactly, without forcing over models of set theory with large cardinals. Magidor[230] then showed that the independence *can* be established by forcing over models with large cardinals in them. This perhaps vindicates Gödel's idea that large cardinals are needed and can be used to "solve" problems about cardinal arithmetic.

---

[226] S. Feferman, in "Reflecting on Incompleteness," *The Journal of Symbolic Logic* 56(1) (1991), pp. 1–49, gives a precise method of measuring the "cost" of accepting consistency statements of increasing strength, in terms of reflection principles.

[227] See e.g. Patrick Dehornoy, Progres récents sur l'hypothèse du continu (d'après Woodin); Seminaire Bourbaki, exposé 915, mars 2003. An English version is available at http://www.math.unicaen.fr/~dehornoy/surveys.html

[228] Easton, William B. "Powers of regular cardinals." *Annals of Mathematical Logic* 1 (1970), pp. 139–178

[229] Devlin, Keith I.; Jensen, R. B. Marginalia to a theorem of Silver. ⊨ISILC Logic Conference (Proc. Internat. Summer Inst. and Logic Colloq., Kiel, 1974), pp. 115–142. Lecture Notes in Math., Vol. 499, Springer, Berlin, 1975.

[230] Magidor, Menachem. "On the singular cardinals problem. II." *Annals of Mathematics*, 2nd series, 106(3) (1977), pp. 517–547.

Nowadays set-theoretical axioms are known which imply SCH. We mention as an example Chang's Conjecture.[231] $(\aleph_{\alpha+1}, \aleph_\alpha) \rightarrow (\aleph_1, \aleph_0)$. A recent result of Matteo Viale[232] shows PFA implies SCH.

9 Concluding Remark: Gödel's Modernism

Recently, Aki Kanamori has pointed out that after Zermelo had clearly separated set theory from logic, Gödel was the one who was prepared to take the linguistic turn and study uninterpreted formal systems from a set-theoretical point of view.[233] This is one sense in which Gödel can be called modern; here is another:

Gödel was nearing the end of his career when generic absoluteness emerged in the 1970s. On his view, reality fixes the intended model of set theory, and we have access to this mathematical reality by intuition. Results about the limits of formalization and formal systems he will therefore ultimately not interpret as revealing limits to the capability of reason to grasp mathematical reality. We saw a particular example of this viewpoint above, in his reaction to Cohen's independence result. On the other hand, formalizations of mathematics will give formal approximations to reason. In that sense, it would be natural for Gödel to see in generic absoluteness a strong justification of his realism and its correlate on the side of the subject, rationalistic optimism. This is primarily because once incompleteness can be explained away as a residual phenomenon, once the statements that we really care about are decided by a theory we in some sense "like," and once the "essential variability in set theory due to forcing" has been explained, then we are on our way to a sufficiently adequate description of the intended model. These results also say something about the robustness of ZFC: namely, it is, after all, a theory which both captures the intuitions of classical mathematicians about sets, and provides a domain for deciding high order questions about sets, in spite of what the incompleteness phenomena may have led people, in the beginning, to believe.

Once again, what is missing in the generic absoluteness approach we have considered here, is, as Shelah points out, to show the uniqueness of forcing. Still, we hoped to draw attention to this sea change in set theory; point out that things are moving along the lines that Gödel anticipated they would. His iron belief in the decidability of the Continuum Hypothesis, radical as it at times seemed, may have been vindicated—by the set theorists.

Acknowledgement We are grateful to the staff of the Department of Rare Books and Special Collections at the Firestone Library of Princeton University, and to Marcia Tucker of the Library

---

[231] Shelah: Cardinal Arithmetic, Oxford logic Guides, vol. 29, Oxford University Press, 1994.
[232] *The Journal of Symbolic Logic* 71(2) (2006), pp. 473–479.
[233] p. 453 of A. Kanamori, "Zermelo and set theory," *The Bulletin of Symbolic Logic* 10(4) (2004), pp. 487–553.

of the Institute for Advanced Study, for facilitating our research, and overall for ensuring such a pleasant stay in the archive; also, we are much obliged to the Institute for Advanced Study which kindly granted permission to quote from Gödel's *Nachlaß*. Many thanks to Piet Hut of the Institute for Advanced Study for inviting Van Atten to participate in the Program in Interdisciplinary Studies which he directs, and to Peter Goddard, the Director of the Institute, for hosting Kennedy, and to the NWO for their support during the final writing of the manuscript.

We wish to thank Joan Bagaria, Leon Horsten, Aki Kanamori, Georg Kreisel, Paolo Mancosu, Göran Sundholm, Boban Velickovic and most especially Jouko Väänänen, for helpful conversation and correspondence, as well as for drawing our attention to some of the sources.

We also thank two anonymous referees for their comments.

Kennedy presented an earlier version of this paper at the workshop "Logicism, Intuitionism and Formalism: What has become of them?", Uppsala, Sweden, August 27–29, 2004. We thank the organizers for the invitation, and the audience for their questions and comments.

Tarski's Practice and Philosophy: Between Formalism and Pragmatism

Hourya Benis Sinaceur

1 Some General Facts About Formalism

1.1 Definitions

The term 'formalism' may have at least three different meanings. First, 'formalism' can be understood as referring to a *mathematical* way of operating. A formalist way of doing mathematics shows how one can get new and innovative results from the mere inspection of symbolic expressions used or coined for mathematical entities or properties. In this wide sense, which is internally connected with a permanent aspect of mathematical practice, one usually speaks of a "formal" rather than of a "formalist" point of view. Leibniz was a great supporter of such a view, promoting symbols and diagrams, be they arithmetical (differential operator, series, determinants) or geometrical (objects of the *analysis situs*), as one way of the "*ars inveniendi*". This point of view is highly represented, from the XIXth century onwards, by the study of mathematical structures defined by axiom systems. Mathematical structuralism aims at more generality, increasing simplicity and unification, deeper understanding and richer fruitfulness. In a second meaning, 'formalism' means a *philosophical* attitude, which seeks an answer not to the question: "how can one do mathematics in a general and very efficient way?" but to the question: "how or on what to ground mathematical practice?" Mathematical structuralism aims at grounding mathematics on the most abstract and general structures, such as those laid for arithmetic or set theory. For instance, Dedekind based the theory of whole numbers on an abstract theory of "simply infinite systems" which presents N as an ordered set satisfying some characteristic conditions. The third and more specific sense of 'formalism' comes from Hilbert's metamathematics, which combines logical analysis of mathematical procedures with philosophical views on the foundations of mathematical practice. This third sense is linked to Hilbert's concern with formal systems of math-

H. Benis Sinaceur (✉)
IHPST (Institut d'Histoire et Philosophie des Sciences et des Techniques),
CNRS-Université Paris 1-ENS
e-mail: sinaceur@canoe.ens.fr

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_16,

ematical theories, to his syntactic study of mathematical proof [*Beweistheorie*], and, notably, to his search for consistency proofs which would secure the soundness of mathematical reasoning against the paradoxes of set theory and would permit to avoid restrictions on classical logic.

1.2 Hilbert's Formalism: Words and Subject. The Paradigm of Algebra

In his essays on the foundations of mathematics, Hilbert did use the German word '*Formalismus*', but *not* to characterize a philosophical attitude towards questions on the nature of mathematical objects or practice. '*Formalismus*' meant 'formal system' or 'formal language', both *technical* concepts of mathematical logic. Sometimes, Hilbert used the word '*Formalismus*' as meaning 'formalization', which is again a *technical* process of mathematical logic.[1] Thus '*Formalismus*' is either the result of a process of formalization or the process itself. Even when Hilbert alluded in his 1931 essay to Brouwer's "reproach of formalism", he took 'Formalismus' *only* in the technical sense and explained that the use of formulas, i.e. formalization, is a necessary tool of logical investigation.

Mostly, '*Formalismus*' is contrasted and correlated with '*Inhaltlichkeit*' or with '*inhaltliche Überlegungen*', and there may be naturally different formalisms or formalized constructions of the same content. The *relationship* between formal processing and informal thinking was nevertheless considered as an *epistemological problem*, just as the consistency problem.[2] And just as for the consistency problem, Hilbert aimed at a logical-mathematical solution, which would make obsolete the epistemological way of questioning and answering. I will briefly sketch this solution below, in 1.4. However, one may note that philosophers of mathematics did not stop until now to be concerned with the relationship between formal setting and content.

Otherwise, Hilbert used the German word '*Formeln*' to speak of mathematical formulas. He distinguished between numerical formulas, such as $2 + 3 = 3 + 2$ or $2 < 3$, and formulas involving variables, namely literal expressions of algebra, such as $a + b = b + a$ or $a < b$. The first ones convey a content which is immediately understandable, while the latter, the "right" formulas, are '*selbständige formale Gebilde*' which have no immediate meaning and are nothing but "objects submitted to the application of our rules".[3] Numerical formulas are formalized through algebraic *formale Gebilde*, which constitute the customary formal part of mathematics. Being entirely determined by definite rules, the formal part of mathematics is con-

[1] Hilbert [29], in Hilbert [36, p. 153]; [30], in Hilbert [36, pp. 165, 170]; [33, pp. 67, 77]; [35, p. 493].

[2] Hilbert [29], in Hilbert [36, p. 153]. As Wolenski suggested to me, it is worth recalling that the contrast between 'form' and 'content' [*Form, Inhalt*] was very popular in Neo-Kantian philosophy, which was very influential at the break of XIX/XX century.

[3] Hilbert [33, p. 72] (my translation).

trollable. Hilbert argued that the "*formaler Standpunkt*", eminently illustrated by algebraic methods, should be expanded to all of mathematics.

> "In algebra we consider the expressions formed with letters to be independent objects in themselves, and the propositions of number theory, which are included in algebra, are formalized by means of them. Where we had numerals, we now have formulas, which themselves are concrete objects that in their turn are considered by our perceptual intuition, and the derivation of one formula from another in accordance with certain rules takes the place of number-theoretic proof based on content.
> Thus algebra already goes considerably beyond contentual number theory."[4]

As a parallel result of this extension, Hilbert upheld a "new formal standpoint",[5] which suited the finitistic building of proof theory. '*Formeln*' came then to be contrasted with the usual mathematical sentences and to designate the corresponding counterpart of the latter in some convenient formalism;[6] they became also "ideal sentences" in a sense analogical to that of Kummer's ideal numbers.

In Hilbert's *early views* the formal standpoint was conceived of as a *conceptual* one and opposed to the algorithmic point of view, supported at that time by Paul Gordan and, to some extent, by Leopold Kronecker. It was then Gordan's calculating methods which were considered as "absolute formalism" in the sense that "formulas were the indispensable supports of the formation of his thoughts, his conclusions and his mode of expression".[7] Hilbert balanced out the exclusive use of symbolic calculations and developed an abstract way of *thinking* and proving that he notably introduced in the theory of algebraic invariants. As we know, Hilbert found out an indirect (through *reductio ad absurdum*) and general (valid for every system of algebraic forms of n variables) proof of the finite basis theorem (1888). Hilbert did not calculate, for each n, the effective number k of the basic invariants, but showed the *existence* of a finite basis for *any* system and for *all* n, by showing that the assumption of the negation of the statement asserting this existence leads to contradictions. Thus, very early in his career, Hilbert advocated the formal point of view first and foremost because it is conducive to general proofs, which make salient structural properties of the problem under consideration. Moreover, the clear distinction between meaning and structure, objects and rules permits to handle uniformly and at once objects of different kinds. Now, the internal efficiency of the formal point of view as well as the applicability of mathematical structures to extra-mathematical phenomena are recognized as valuable. But, from the philosophical standpoint, what is at stake in axiomatic definitions and in structural existence proofs is the meaning of mathematical existence. Does existence follow from the supposed compatibility of some selected axioms as long as no contradiction appears in their consequences?

---

4 Hilbert [33, pp. 71–72]; English translation in van Heijenoort [76, p. 469]. The standpoint of formal algebra is presented in a different way in Hilbert and Bernays [37, pp. 29–32]: the elementary algebra, defined as the elementary theory of rational functions with integer coefficients, is included in the domain of elementary contentual inference.

5 Hilbert [30], in Hilbert [36, p. 168].

6 Hilbert [30, p. 174]; [31, p. 179]; [32, p. 175]; [33, p. 66].

7 Reid [50, p. 30].

Or are existential statements "empty inventions of logicians"[8] as long as we don't have an actual realization?

1.3 Brouwer's Criticism

Brouwer, who stood up for the second opinion, was *the one* who used for the first time the word 'formalism' as denoting the first opinion. In a 1909 review of Mannoury's *Methodologisches und Philosophisches zur Elementar-Mathematik*, Brouwer wrote that "the formalist conception recognizes no other mathematics than the mathematical language and it considers it essential to draw up definitions and axioms and to deduce from these other propositions by means of logical principles which are also explicitly formulated beforehand. This has two consequences . . ., namely the priority of infinite over finite numbers and the belief in higher cardinalities than that of the continuum".[9] In his famous 1912 essay, Brouwer added other considerations, the analysis of which shows that he took formalism as purely and simply antithetic to his intuitionism.[10] By the label 'formalism', Brouwer referred to a global *philosophical* attitude involved in classical methods of analysis as well as in set theory and in modern axiomatic theories, which use the language and the means of symbolic logic. According to Brouwer [9], formalism encompasses three main assumptions. First, formalism admits the existence of an entity on the grounds of its supposed non-contradictory definition. Such an existence, says Brouwer, is merely a linguistic existence, which corresponds to the method of posing meaningless axioms and deducing from meaningless relations some other meaningless relations in the language of symbolic logic. The second point is a consequence of the former: being meaningless, formalist assertions miss intuitive thinking and put forward logical support for self-evident principles, such as the principle of complete induction. In particular, the aim at consistency-proofs is anchored in a logical, i.e. a *non-mathematical, conviction* of legitimacy. Using the term 'conviction', which is Brouwer's word,[11] means that even logical procedures may be rooted in (or supported by) a subjective belief. That is a harsh criticism against the supposed absolute objectivity of logic, which formalists put in contrast with subjective intuition. Moreover, and more seriously, the aim at consistency-proofs leads to a vicious circle, as Poincaré [49] already pointed out. Last but not least, the third point highlighted by Brouwer is the Platonist assumption of a universe of mathematical entities,

[8] H. Weyl [80], English translation in Mancosu [44, p. 133].

[9] Brouwer [14, p. 121].

[10] Brouwer [9], in Brouwer [14, pp. 123–137].

[11] Brouwer [14, p. 125]: "It is true that from certain relations among mathematical entities, which we assume as axioms, we deduce other relations according to fixed laws, in the conviction that in this way we derive truths from truths by logical reasoning, but this non-mathematical conviction of truth or legitimacy *has no exactness* whatever and is nothing but a vague sensation of delight arising from the knowledge of the efficacy of the projection into nature of these relations and laws of reasoning" (my emphasis).

subsisting independently of our thought and, so to speak, ready to be structured according to the laws of classical logic and set theory.

Brouwer took just the opposite views on the three points: he advocated intuition as the original *material source* and *justification* of mathematical practice; he argued that the language is a "non-mathematical auxiliary" for helping memory or conveying communication (along with misunderstanding); he rejected the Platonist static view, defending a dynamic conception in which an entity exists for a mathematician if it is actually constructed by some effective process. According to Brouwer, formalist purported foundations of mathematical laws on the axiomatic method are nothing but mere linguistic explanations, devoid of content, which, as such, don't really (materially) explain anything. By contrast, intuitionism explains the accuracy of mathematical reasoning by the material self-development of human mind from one original insight. The original insight [*Urintuition*] is the *a priori* insight of time, from which is derived, by "abstraction",[12] the first mathematical insight, namely the intuition of the number 2 and, step by step, the intuition of whole numbers and of any mathematical construct grounded on whole numbers. The intuition of two-ity is the fundamental phenomenon of mathematical thinking.

It is worth noting that Brouwer acknowledged the mediating role of mathematical abstraction in the very first intuitive process. Mathematical abstraction produces indeed the very first empty form, which constitutes the first *substratum*, "as basic intuition". That is to say that Brouwer did not oppose intuition to form in the *substantial* mathematical process. It is quite the contrary, as it is clear from the passages quoted in footnote 12. Brouwer did naturally not reject the *formal way of practice*. What he rejected was locating the *justification* of mathematical substance in symbolic schemas and formal deductions, which are, according to him, only an *external dressing*. Brouwer rejected also formalism, not as a mathematical way, but as *philosophy*, or, more accurately as mathematical project to solve philosophical

[12] See Brouwer [11], in Brouwer [14, pp. 418–419], English translation in Mancosu [44, p. 46]: "Mathematical action can only reach its full development at the higher stages of civilization when *mathematical abstraction* comes into play. By means of mathematical abstraction man strips two-ity of its material content and retains it as an empty form, the common substratum of all two-ities. This common substratum of all two-ities forms the *Primordial Intuition of Mathematics* (*die Urintuition der Mathematik*), which in its self-unfolding also introduces the infinite as a thought-reality and produces the collection of natural bumbers..., as well as the real numbers, and finally the whole of pure mathematics" (Brouwer's emphasis).

See also Brouwer [12], in Brouwer [14, p. 482]: "Mathematics comes into being, when the two-ity created by a move of time is divested of all quality by the subject, and when the remaining empty form of the common substratum of all two-ities, as basic intuition of mathematics, is left to an unlimited unfolding, creating new mathematical entities in the shape of *predeterminately or more or less freely proceeding infinite sequences* of mathematical entities previously acquired, and in the shape of *mathematical species* i.e. properties supposable for mathematical entities previously acquired and satisfying the condition that if they are realized for a certain mathematical entity, they are also realized for all mathematical entities which have been defined equal to it" (Brouwer's emphasis).

problems.[13] He especially disputed the prominent role that formalists, as well as logicists, gave to logic in the foundations of mathematics. He summed up the debate between formalism and intuitionism in the following ironic words: "The question where mathematical exactness does exist, is answered differently by the two sides; the intuitionist says: in the human intellect, the formalist says: on paper."[14]

Three observations need to be made. Firstly, just as it is a mistake to believe that, in Brouwer's mind, intuition excludes abstraction, it would be wrong to trust the popular image of formalism and to believe that for "formalist" mathematicians intuition plays no role at all. From the structural point of view, it is for the sake of a flawless rigor that intuition must be submitted to a logical and axiomatic analysis, as, for instance, Dedekind did for Number theory[15] and Hilbert for Euclidean geometry.[16] Intuition is admitted as giving the matter to be analyzed, criticized, and generalized, but this chronological priority does not legitimate an *ontological* or *epistemological* primacy.

Second remark. The belief in a pre-existent universe of mathematical objects characterizes more sharply logicism than formalism. Now in his 1912 essay, Brouwer did not even mention logicism as a separate point of view. According to the title, he distinguishes only two contrary options: formalism and intuitionism, which, he thinks, cannot understand each other, because "they do not speak the same language". And indeed, grounding both on the laws of classical logic, and in particular on the principle of excluded middle, logicism and formalism speak, in Brouwer's opinion, the same language. In the 1909 review Brouwer mentioned among the formalists Dedekind, Peano, Russell, Hilbert and Zermelo. In a later paper he added Cantor and Couturat to the list:

> "... the *Old Formalist School* (Dedekind, Cantor, Peano, Hilbert, Russell, Zermelo, Couturat), for the purpose of a rigorous treatment of mathematics *and logic* (though not for the

---

[13] Kreisel [42, p. 158], observed similarly that "the real opposition between Brouwer's and Hilbert's approach was not at all between formalism and intuitive mathematics, but between the conception of what constitutes a foundation".

[14] Brouwer [14, p. 125].

[15] See [17, pp. 99–100]: "How did my essay [*Was sind und was sollen die Zahlen?*] come to be written? Certainly not in one day; rather it is a synthesis constructed after protracted labor, based upon a prior analysis of the sequence of natural numbers just as it presents itself, in experience, so to speak, for our consideration. What are the mutually independent properties of the sequence N, that is, those properties that are not derivable from one another but from which all others follow? And how should we divest these properties of their specifically arithmetic character so that they are subsumed under more general notions and under activities of the understanding *without* which no thinking is possible at all but *with* which a foundation is provided for the reliability and completeness of proofs and for the construction of consistent notions and definitions?".

[16] See Hilbert's [27, p. 1]: "Die Aufstellung der Axiome der Geometrie und die Erforschung ihres Zusammenhanges ist eine Aufgabe, die seit Euklid in zahlreichen Abhandlungen der mathematischen Literatur sich erörtet findet. Die bezeichnete Aufgabe läuft auf die *logische Analyse* unserer räumlichen *Anschauung* hinaus" (my emphasis). See Webb's valuable comments on Hilbert's geometrical methods, Webb [77, Chapter III]: in short, Hilbert did not eschew space intuition, he made the axioms of geometry more explicit "in order to determine both explicit and implicit uses of space intuition" (p. 110).

> purpose of choosing the subjects of investigation of theses sciences) rejected any element extraneous to language and logic."[17]

It is clear that Brouwer made no difference between logicists and formalists. There were two major reasons for gathering them under the same label. First, logicists and formalists both distrusted intuition as being an unreliable access to mathematical objects and a shaky ground for mathematical practice. Second, they both developed projects that were intended to ground mathematics on logic, logic being understood as yielding schemas of correct derivation for a formal theory, especially that of natural numbers, the base of the rest. Hilbert's aim was to establish a simultaneous foundation of the laws of arithmetic and logic.[18] But, while logicists believed that mathematical entities were *discovered* by purely logical thought, formalists advocated explicitly the *free creation* or construction of new *concepts* [*Begriffsbildungen*],[19] even of old familiar notions such as that of whole numbers. For Dedekind indeed numbers are "free creations of the human mind", which "serve as a means of apprehending more easily and more sharply the difference of things".[20] They are also *objective instruments* for grasping the multiplicity. In a similar spirit, and as a justification for the transfinite numbers, Cantor argued that "the human mind has an unlimited ability to progressively construct classes of numbers ... with increasing *powers* [*Mächtigkeiten*]".[21] Here again, the mathematical universe (including infinite sets) is originating from the human mind. Hilbert supported this view: in his opinion the theory of transfinite numbers was "the most admirable flower of the mathematical intellect and in general one of the highest achievements of purely rational human activity".[22]

Rigorously speaking, this conception of a creative mind would entail a philosophical subjectivism, i.e. the conception of a *subjective* existence of those created concepts. Now, "subjective existence" might mean existence *in* our mind, or existence *dependent* of our mind. Formalists generally choose the second (weaker) meaning while Brouwer assumed the first too.[23]

---

[17] Brouwer [13, p. 508] (Brouwer's emphasis).

[18] Sieg [54] showed how Hilbert moved progressively from "a critical logicism through a radical constructivism toward finitism".

[19] Typical expression of Hilbert's style. See, for instance [28, p. 183]; [32, p. 170]; [33, p. 65] (translated by [ways of] "forming notions" and by "mathematical definitions" in van Heijenoort [76], respectively on p. 376 and p. 464; the literal translation: "concept-formations" of Mancosu [44, p. 189], seems preferable to me).

[20] Grounding on the text quoted in footnote 15, one must precise that what Dedekind considered as "a free creation of the human mind" were not the familiar numbers of our naive arithmetical experience, but the "shadowy forms" that Dedekind was making free from any particular content and which "are always the same in all ordered simply infinite systems, whatever names may happen to be given to the individual elements" (Dedekind [16], Definition 73).

[21] Cantor [15, p. 177] (Cantor's emphasis).

[22] Hilbert [32, p. 167], in van Heijenoort [76, p. 373].

[23] See the passages quoted above in footnote 12 and the following excerpt: "The fullest constructional beauty is the *introspective beauty of mathematics*, where instead of elements of playful causal acting, the basic intuition of mathematics is left to free unfolding. This unfolding is *not*

Dedekind and Hilbert rejected no less vigorously than the logicists (Bolzano or Frege) subjectivism as meaning existence *only in* our mind.[24] They conceived of axiomatic definitions as objective structural laws of mathematical processes *in concordance with the laws of thought*. Such a conception fits the Kantian view according to which the human mind [*Verstand*] is entitled with the legal power of organizing experience: mathematical concepts depend on the structure of the human mind *and* they help to organize the phenomenal world.[25] But, while focusing on what they accepted as laws of mind, formalists generally did not accept the *apriority* of space and time as the formal setting making experience possible through the application of categories. – On his side, Brouwer abandoned the apriority of space but advocated resolutely the apriority of time. – Formalists admitted at the same time that mathematical concepts were created (constructed) rather than discovered and that the construction was neither arbitrary nor conventional but corresponded to some objective phenomenal connections. I would say that the creationist view was bound with the assumption of the objective adequacy of mind with the physical world. Such an adequacy, if it holds, leads to assume a kind of *immediate* consistency of the created concepts, which become questionable only when some contradiction appears in their consequences.

Third remark. Denying a foundational status to intuition, as Hilbert did firstly, is hardly compatible with a coherent and strict Platonism (such as that one defended by Gödel). – But associating anti-Platonism with foundational intuition, as Brouwer did, is not less problematic, unless intuition does not mean intuition *of* something exterior to the mind and reduces to mere introspection. An implication accepted by Brouwer, as I recalled right above. – In fact, matters were (and are) really complicated and the difficulties inherent to connecting a one-sided and clear-cut philosophical attitude with the multi-faceted mathematical practice have been explained in Bernays' famous essay on Platonism in mathematics.[26] Developing a criticist remark passed by Hilbert on Frege's "extreme conceptual realism",[27] Bernays distinguished two kinds of Platonism: (1) the restricted one, which considers abstract entities as nothing but a sort of "ideal projection of a domain of thought" (a precise explanation of the meaning of this expression would lead to some difficulties, that we do not want to address in this paper); (2) the extreme Platonism in the sense of a conceptual realism, which postulates an independent world of ideas containing all

---

bound to the exterior world, and thereby to finiteness and responsibility", Brouwer [14, p. 484] (my emphasis).

[24] Hilbert [33, p. 80], in van Heijenoort [76, p. 475]:"it is part of the task of science to liberate us from arbitrariness, sentiment, and habit and to protect us from the subjectivism that already made itself felt in Kronecker's views and, it seems to me, finds its culmination in intuitionism".

[25] Such a view leads often to some kind of instrumentalism. Since formalists generally reject the idea of an ontological foundation for mathematics, they tend to support a positivistic justification, according to which mathematical methods are *epistemological tools* in coping with the empirical world.

[26] Bernays [7]. Reprint in P. Bernays, *Philosophie des mathématiques*, Paris, Vrin, 2003, pp. 83–104.

[27] Hilbert [30, p. 162].

the objects and relations of mathematics. According to Bernays, Russell's antinomy ruined only the extreme Platonism (which was supported by Bolzano and Frege, and not by Dedekind or Hilbert). The minimal assumption of a restricted Platonism is to admit the set of natural numbers. Bernays observed that for some theories even this minimal assumption is not necessary: Kronecker introduced algebraic numbers without supposing the totality of whole numbers. But for other domains, such as infinitesimal analysis or function theory, the minimal assumption is needed. The strongest assumptions of Platonism are made in Cantorian set theory.

The fact is that, in his 1912 paper, Brouwer explicitly aimed to challenge the validity of the axioms of set theory stated by Zermelo in 1908. It was therefore natural that Brouwer associated Platonism, a kind of which supported, at least *tacitly*, the underlying universe of sets, with the general formal point of view. Now, working with actual infinite sets does not necessarily means that one believes that they exist prior to and independently of their being thought. A formalist, even if he is a set-theorist, need not to support an extreme Platonist view of pre-existing sets; he may content himself with some restricted view. *But certainly*, applying to infinite sets the principle of excluded middle is rightfully questionable in any case. Brouwer did not reject the infinite. He simply understood it as a "thought-reality" (se above, footnote 12) and he did not accept higher cardinalities than those of natural numbers and real numbers. But, he definitely rejected the general use of the principle of excluded middle, which has been classically used by mathematicians for centuries, and he replaced the static "spatial" conception of sets involved in it with a dynamic self-unfolding of *spreads* based on the apriority of time.

Although Brouwer's [9] paper did not make explicit reference to Hilbert, the attack against Hilbert's [28] paper on the foundations of logic and arithmetic was very clear. In particular, Brouwer repeated Poincaré's devastating argument [49] against the admission of the principle of complete induction *as an axiom*, instead of accepting it *as intuitively evident*.

1.4 Hilbert's Defense of Formalism

As is well known, Hilbert took Poincaré's and Brouwer's objections seriously and he associated the latter with Kronecker's reductionism to whole numbers. From 1918 to 1931, he published a series of essays, in which he introduced a "new mathematics", namely metamathematics, and developed technically and philosophically his famous finitistic program. It is not my purpose to enter in the technical details of this program[28] and its subsequent reorientations. I would like only to point out some philosophical modifications it involved.[29] Hilbert supported indeed a *new* for-

[28] See in particular the recent paper by R. Zach [86] on ε-calculus and consistency proofs in Hilbert's school.

[29] See Benis Sinaceur [3], and Sieg [54] for a thorough analysis of Hilbert's unpublished notes of lecture courses from 1917 to 1922.

mal point of view, which incorporated what he called "the constructivity principle" and some other intuitionistic insights in a much more systematic and radical "formalism" than that [28] which aroused Brouwer's polemic notion of formalism. Brouwer's criticism acted in a performative way and pushed Hilbert to present logical inferences as "purely formal operations with letters"[30] and to play fully the formula game in a constructive way.

a) Hilbert was urged by Brouwer's and Weyl's objections to make precise the concept of formal system through a kind of *material* implementation. He considered that one must have something primitive and irreducible to begin with. He then changed his mind about intuition and logic and accepted to give intuition a basic role in the formal treatment. From 1922 onwards, he gave up Frege's and Dedekind's idea to provide for arithmetic a foundation that would be independent of all intuition and experience and he claimed that "as a condition for the use of logical inference and the performance of logical operations, something must already be given to our faculty of representation, certain extra-logical concrete objects that are intuitively present as immediate experience prior to all thought".[31] Thus, Hilbert admitted that the mathematician starts with an intuitive notion of natural numbers, what was Kronecker's, Poincaré's and Brouwer's common claim. However, what he regarded as intuitive was not a familiar or naïve notion but a finite stock of symbols given to spatial perception and having, in themselves, no meaning at all. The objects of (formal) arithmetic are not numbers but numerals, mere shapes or types of the actual signs written down on a sheet of paper. The sign '1' is a number as well as any finite sequence beginning and ending with 1 provided that the sign '+' is placed between two successive 1. Thus, instead stating by an existence axiom that "each number has a successor", Hilbert introduced a progressive construction. Answering Poincaré's objection Hilbert distinguishes this combinational way to construct finite numbers as numerals from the principle of complete induction; the latter is a formal principle based on the induction axiom, which uses the general concept of whole number, while the former is a contentual [*inhaltlich*] composition. Hilbert maintained however that the formal principle, along with the other axioms of his formal system for natural numbers, has to be justified by a consistency proof.[32]

---

[30] Sieg [54, p. 9]. Sieg notes that this presentation requires "a formal language (for capturing the logical form of informal statements), the use of a formal calculus (for representing the structure of logical arguments), and the formulation of 'logical' principles (for defining mathematical objects)". Sieg highlights Hilbert's and Bernay's contribution to the creation of modern mathematical logic (pp. 11–12).

[31] Hilbert [30, p. 162]; [32], in van Heijenoort [76, p. 376]; [33], in van Heijenoort [76, p. 464].

[32] Weyl [81] gave right to Poincaré: even if a consistency proof could justify the formal principle, it would not justify the intuitive one. Therefore one need not express mathematical induction as an axiom; one may just make its self-evidence and primarity explicit and accept it as a characteristic mark of contentual mathematical thought [76, p. 483]. In a similar way, Brouwer [10] pointed out again the circularity of the endeavor of justifying the formal proposition by a consistency proof, "since this justification rests upon the (contentual) correctness of the proposition that from the

The trick of the new formal point of view was to apply to mere types of signs a contentual constructive process and, thus, to reverse the traditional relationship between formal and content: the perceptible object is formal and it is submitted to a contentual process. Mathematical thoughts, in the customary sense, are mirrored by concretely exhibited formulas, which are either primitive sentences or sentences provable, at some stage, from those primitive ones, and the whole of mathematics is duplicated by a stock of formulas. Besides mathematical signs, those formulas contain logical signs, which are, too, divested of all meaning. In turn, proofs are indeed perceptible arrays or sequences of formulas, which concretely present the formal images of customary mathematical inference so that "contentual inference is replaced by *manipulation* of signs according to rules".[33] Thus, Hilbert confirmed Brouwer's account of formalism[34] and went even further: he understood insight as physical perception and formalism as a mechanistic operating with mere signs, formulas and arrays. The latter are indeed, according to his new point of view, the concrete and surveyable objects of metamathematics, which is "the contentual theory of formalized proofs"[35] and which would use only contentual arguments for establishing the consistency of the formalized system of arithmetic. The contentual character ultimately rests, on the one hand, upon Hilbert's conviction that metamathematical induction, operating on finite existing totalities, was contentual,[36] just as intuitive composition and decomposition of numerals, and, on the other hand, upon the fact that the consistency proof amounts to show that one cannot derive the formula $0 \neq 0$ in the system under consideration, "a task that fundamentally lies within the province of intuition".[37] Hilbert wanted to renounce neither Cantor's paradise nor Aristotle's laws of logic. He aimed at justifying them contentually and by finitistic, i.e. strictly constructive, means.[38] To restore the security shaken by

---

consistency of a proposition the correctness of the propoeition follows, that is, upon the (contentual) correctness of the principle of excluded middle" [76, p. 491].

33 Hilbert [32], in van Heijenoort [76, p. 381] (my emphasis).

34 See Brouwer [10]: "the differentiation between a construction of the 'inventory of mathematical formulas' and an intuitive (contentual) theory of the laws of this construction″ . . .penetrated into the formalistic literature with Hilbert [30]". Brouwer mentioned that he spoke with Hilbert on that issue in the autumn of 1909 and hence he did not appreciate naming 'metamathematics', without "observing proper mention of authorship", what was, according to him, his notion of 'mathematics of the second order'.

35 Hilbert [31, p. 181] and Hilbert [32], in van Heijenoort [76, p. 385].

36 See the comments on Hilbert's metamathematical induction in the introductory note to Weyl's 1927 paper in van Heijenoort [76, pp. 480–482].

37 Hilbert [33], in van Heijenoort [76, p. 471].

38 Hilbert never spelled out the exact boundaries of finitistic means. However, Hilbert [31] mentioned explicitly induction and recursion on existing finite totalities. Hilbert [32] (in [76, pp. 377–378] explained how to prove that there exist infinitely many primes by proving first the partial proposition: for a fixed prime p there exists a prime q such that $p < q \leqslant p! + 1$. In the latter proposition the existential quantifier is bounded (applied to a finite totality) and can be replaced by a finite disjunction. Hilbert's device here is similar to Skolem's method of restricted domains of existence (Skolem [55], in [76, pp. 302–333]). See Tait [57]. Sieg [54, pp. 28–29] shows that Hilbert's

the paradoxes and the attacks against the actual infinite, Hilbert saw no other way than a finitary consistency proof of the "ideal" picture of mathematics he constructed step by step.

b) Hilbert thought that it was necessary to consider the formal picture of customary mathematics. The reason is the following:

> "... **even elementary mathematics** contains, first formulas to which correspond contentual communications of finitary propositions (mainly numerical equations or inequalities, or more complex communications composed of these) and which we may call the *real propositions* of the theory, and, second, formulas that – just like the numerals of contentual number theory – in themselves mean nothing but are merely things that are governed by our rules and must be regarded as the *ideal objects* of the theory."[39]

Therefore, Hilbert fully assumed the formula game. He maintained that

> "the formula game enables us to express the entire thought-content of the science of mathematics in a uniform manner and develop it in such a way that, at the same time, the interconnections between the individual propositions and facts become clear. To make it a universal requirement that each individual formula then be interpretable by itself is by no means reasonable; on the contrary, a theory by its very nature is such that we do not need to fall back upon intuition or meaning in the midst of some argument".[40]

Moreover, Hilbert credited the formula game with a philosophical significance; he claimed that it expressed the "technique of our thinking". According to Hilbert, his proof theory provided "a protocol of the rules according to which our thinking actually proceeds".[41] This is clearly a mechanistic view of mathematical thought.[42]

---

idea is "strikingly similar to Weyl's viewpoint" in Weyl [79]. On his side, Zach [86] establishes (p. 220) that the general schema of primitive recursion was already mentioned in Hilbert's unpublished course of 1921–1923. Moreover, he argues that Hilbert's outlook was "markedly different" from Skolem's [55] (suggesting that there was no influence either way). Third, he challenges the generally admitted thesis, according which 'finitistic' means 'primitive recursive', stressing that Hilbert considered Ackermann's 1924 proof to be finitistic, although this proof used transfinite induction up to $\omega^{\omega^{\omega}}$ (I thank P. Mancosu for drawing my attention to Zach's paper).

[39] Hilbert [33], in van Heijenoort [76, p. 470] (Hilbet's emphasis, bold types are mine). Sieg [54] throws new light on this point. He quotes the following passage from Hilbert's notes for the winter term 1920: "We have to extend the domain of objects to be considered; i.e. we have to apply our intuitive considerations also to figures that are not number signs″ ... "the figures we take as objects must be completely surveyable and only discrete determinations are to be considered for them. It is only under these conditions that our claims and considerations have the same reliability and evidence as in intuitive number theory".

[40] Hilbert [33], in van Heijenoort [76, p. 475].

[41] Hilbert [33], in van Heijenoort [76, p. 475].

[42] Contrast with Heyting [26], in Mancosu [44, p. 311]: "every language, including the formalistic one, is only a tool for communication. It is in principle impossible to set up a system of formulas which would be equivalent to intuitionistic mathematics, for the possibilities of thought cannot be reduced to a finite number of rules set up in advance".

c) Hilbert did acknowledge that the validity of the principle of excluded middle was contentually limited to finite sets,[43] but he sought the means to legitimately extend it to the transfinite. For this purpose he introduced the logical "transfinite axiom" by means of the tau or epsilon-function so that he could introduce the quantifiers and derive the principle of excluded middle. Thus, he used the epsilon-function to carry out pure existence proofs that he advocated once more, insisting on the brevity and the economy of thought they allow. Moreover, Hilbert noted that even if one were not satisfied with consistency, which actually constituted the core of his proof theory, one had to acknowledge the significance of the consistency proof as a general method of obtaining from general proofs finitary proofs carried out by means of the epsilon-function.[44] This perspicuous remark inspired later a whole trend of proof-theoretic work, notably illustrated by some known papers of G. Kreisel.[45]

Concluding this rough sketch, I have to stress that I was concerned here only with aspects of Hilbert's work which may illustrate the formalist view he supposedly championed. I did not aim at supporting Brouwer's opposition to Hilbert, but at understanding what Brouwer meant by 'formalism' and to what extent Hilbert's methods and reflections matched the label Brouwer created. However, not only Hilbert's achievements transcended the boundaries of this label in many respects, but also the notion of formalism evolved so much as to *not* coincide at all with Brouwer's description.

2 Tarski's Semantic Formalism

2.1 Metamathematics Reoriented

Although he borrowed and transformed many technical elements and some views from each of the three standpoints: logicism, formalism and intuitionism, Tarski supported explicitly and exclusively the philosophy of none of them. Moreover, he repeatedly claimed he could develop his mathematical and logical investigations without reference to any particular philosophical view concerning the foundations of mathematics. He was eager to disconnect his results from any definite philosophical view, as well as from his personal (varied and variable) leanings. He believed that scientific precision was inversely proportional to philosophical interest, even though he had strong interest in philosophical issues.

[43] Hilbert [31], in Hilbert [36, pp. 181–182]. See Brouwer's comment in Brouwer [10].

[44] Hilbert [33], in van Heijenoort [76, p. 474].

[45] For instance Kreisel [40, p. 156]; Kreisel [41, pp. 361–362]; Kreisel [42, p. 162]: "As far as piecemeal understanding is concerned, its [Hilbert programme] importance consists of having led to the fruitful study of the constructive aspects of axiomatic systems . . . My own interest . . . does not go one way, i.e. the elimination of non-constructive methods, but I find that greater facility with non-constructive methods comes from a study of their constructive aspects".

We are in front of a new fact in the history of modern mathematical logic: the non-tacit and expressively assumed splitting between logical work as such, on the one hand, and, on the other hand, assumptions or beliefs about the effective or legitimate ways of doing that work and about the nature of the mathematical and logical entities linked with those ways.

Russell aimed to make philosophy as accurate as mathematics. Hilbert aimed to substitute mathematics to philosophy for tackling some important questions falling within the theory of mathematical knowledge – this was the epistemological aim of his metamathematics, which led him to the technicalities of his syntactic study of proof.[46] Tarski wanted to separate logical results from ontological and epistemological problems of the foundations of mathematics, so that those results become easily understandable and usable by working mathematicians. He did not take sides in the fight about how to get mathematical entities well grounded and mathematical practice rightly justified. He was fighting for a *new place* for logic within mathematics, showing how to use fruitfully logical tools in the mathematical research. Solomon Feferman, who studied with Tarski at Berkeley from 1948 to 1957, testified that Tarski did have a very strong motivation, not only to make logic mathematical (Hilbert had the same aim, and before many logicians as well), "but also and at the same time to make it of interest to mathematicians".[47] This is why Tarski objected to *restricting* the role of logic to the foundations of mathematics. He always kept taking his initial aim, which was to make metamathematics a full mathematical field in its own right, like *any other* mathematical discipline, such as arithmetic or geometry. He claimed in a 1930 paper that "formalized deductive disciplines form the field of research of metamathematics roughly in the same sense in which spatial entities form the field of research in geometry".[48] This claim of constituting metamathematics as a mathematical discipline was not fundamentally different from Hilbert's viewing *Beweistheorie* as a "new mathematics". And we may add that, in some respect, Tarski agreed with Hilbert's positivist claim, according to which "mathematics is a science without [philosophical] assumptions".[49] But while Hilbert kept investigating mathematical-logical foundations, in order to eradicate philosophical dogmatism and, eventually, to interpret Kant's *a priori* as the finite mode of thought,[50] Tarski did not think he was (only) contributing to the

---

[46] Hilbert [32, p. 180]; in van Heijenoort [76, pp. 383–384]: "our proof theory ... is not only able to secure the foundations of the science of mathematics; I believe, rather, that it also opens up a path that ... will enable us to deal for the first time with general problems with fundamental character that fall within the domain of mathematics but formerly could not even be approached." See also Bernays' comment: "The great advantage of Hilbert's procedure rests precisely on the fact that the problems and difficulties that present themselves in the grounding of mathematics are transferred from the epistemological-philosophical domain into the domain of what is properly mathematical. Mathematics here creates a court of arbitration for itself, before which all fundamental questions can be settled in a specifically mathematical way...″, Bernays [5], in Mancosu [44, pp. 221–222].

[47] See Duren [19, p. 402].

[48] Tarski [58], in Tarski [70, **I**, p. 313].

[49] Hilbert [33 p. 85].

[50] Hilbert [34], in Hilbert [36, pp. 383–385].

foundations of mathematics. He thought he was building a new mathematical branch on its own. Let us note, in passing, that the implicit epistemological attitude behind this thought was squarely opposed to the intuitionistic view according to which logic is extraneous to mathematical substance. The success of Tarski's enterprise came neither rapidly nor obviously. Still in 1955 Tarski was insisting on the bridge to be built or reinforced, in order to bring mathematicians close to logical methods. He and Leon Henkin wrote to E. Hevitt a letter (September 26, 1955) for supporting the idea of a summer institute on logic at Cornell University; they argued as follows:

> "There are some mathematicians who are not familiar with the many directions in which this field [of logic] has recently developed. These mathematicians have the feeling that logic is concerned exclusively with those foundation problems which originally gave impetus to the subject; they feel that logic is isolated from the main body of mathematics, perhaps even classify it as principally philosophical in character. Actually such judgments are quite mistaken. Mathematical logic has evolved quite far, and in many ways, from its original form. There is an increasing tendency for the subject to make contact with other branches of mathematics, both as the subject and method."[51]

Indeed, Tarski strove to give logic a *heuristic* role in the growth of mathematical theories. As I pointed it out elsewhere,[52] Tarski had no scruples about using formal methods and expanding them from mathematics to mathematical logic. It was he who initiated, in the 1930s, the *heuristic shift* in modern logic. A long time after the beginnings of this shift, Georg Kreisel commented as follows: "the passage *from* the foundational aims for which various branches of modern logic were originally developed *to* the discovery of areas and problems for which logical methods are effective tools ... did not consist of successive refinements ... but required radical changes of direction".[53]

Thus, the heuristic shift reoriented the direction of foundational studies, breaking the hope that the latter would yield a final guarantee [*Sicherung*] of mathematical reasoning. Tarski thought that the aim to provide for mathematicians "a feeling of absolute security" was "far beyond the reach of any human science"; it pertained to "a kind of theology".[54] Therefore, the non-theological aim of metamathematics was not to secure mathematics but to develop it. Tarski showed through some very significant examples, especially that of definable sets of real numbers, that metamathematics is nothing but just a new branch of "ordinary" mathematics.[55] He stressed many times the following opinion:

> "The distinction between mathematics and metamathematics is rather unimportant. For metamathematics is itself a deductive theory and hence, from a certain point of view, a part

[51] Tarski's papers, Bancroft Library, quoted by Joseph W. Dauben [18, p. 233].
[52] Benis Sinaceur [2, Part IV] and [4, pp. 56–57].
[53] Kreisel [43, p. 139] (Kreisel's emphasis).
[54] Tarski [73, p. 160].
[55] It is today well known that the basic concept of real algebraic geometry, i.e. the concept of semi-algebraic sets originates, conceptually if not through actual historical development, in Tarski's concept of definable sets of real numbers.

> of mathematics... Also from a practical point of view, there is no clear-cut line between metamathematics and mathematics proper".[56]

Tarski rejected also the clear-cut border that Hilbert put between the two connected fields, in order to neutralize Poincaré's criticism.[57] But bringing metamathematics near to mathematics is bringing it far from philosophy. After Tarski, I will therefore distinguish the logic-mathematical level from the philosophical one. I propose to consider first Tarski's formalism in his mathematical and metamathematical *practice*, and to leave for a third part of this paper Tarski's *philosophical* considerations.

2.2 Tarski's Version of Formalism

To begin with, one must again highlight one significant fact. From the start of his career, Tarski was combining different technical ways which might have been judged previously incompatible.

I have noted this multi-sided methodology a long time ago.[58] J. Wolenski explains it as a consequence of the philosophical liberalism and the scientific ideology of the Warsaw School of logic.

> "Since the school did not consider itself restricted by any philosophical assumption, it could freely observe the principle of 'logic for logic's sake' and take up, without any a priori prejudices, all those investigations that were interesting from the logical point of view".[59]

However, the general spirit of Tarski's logico-mathematical work was formalist, in a sense I shall explain right now.

a) First of all, Tarski adopted axiomatics and Hilbert's metamathematics, word and concept. However, the issue at stake was for him not only the structure of mathematical proof in a formal system, but rather the structure of the *deductive theories*[60] themselves, with a special eye on the most ancient and daily practiced mathematical domains, such as Euclidean geometry and real numbers. Tarski

[56] Tarski [66], in Tarski [70, **II**, p. 693].
[57] Hilbert [30, p. 165]: Hilbert explained that he would develop a standpoint which makes possible "a strong and systematic separation, in mathematics, between formulas and formal proofs on the one hand and, on the other hand, contentual considerations". Herbrand believed that the very strict distinction between mathematics and metamathematics would put an end to discussions on the foundations of mathematics, Herbrand [38, p. 39].
[58] H. Benis Sinaceur [2, Part IV].
[59] J. Wolenski [82, p. 192].
[60] Tarski distinguished between deductive systems and deductive theories. See, for instance, Tarski [61], in Tarski [69, p. 343, footnote 1]: "By deductive theories I understand here the *models* (realizations) of the axiom system which is given in Section 1.... On the other hand, deductive systems (in the domain of a particular deductive theory) are certain special sets of expressions which I shall characterize at the beginning of 1 as well as in Definition 5 of Section 2".

widened the scope of metamathematics, which no longer coincided with proof theory and the search for finitary consistency proofs. In his practice, he did not hesitate to use infinitistic and impredicative methods and he admitted first-order logic with infinitely long expressions, even though he actively participated in the early forties (with Carnap and Quine) to the endeavour to construct a finitistic language for science.[61] Retrospectively, Tarski noted:

> "As an essential contribution of the Polish school to the development of metamathematics one can regard the fact that from the very beginning it admitted into metamathematics *all fruitful methods*, whether finitary or not. Restrictions to finitary methods seem natural in certain parts of metamathematics, in particular in the discussion of consistency problems, though even here these methods may be inadequate. At present time it seem certain, however, that exclusive adherence to these methods would prove a great handicap in the development of metamathematics".[62]

b) Second, studying some deductive theory, Tarski paid attention to all the possible meanings of its axioms system. Confirming Lesniewski's idea, and therefore in connection with Husserl's phenomenology and the Vienna semantic tradition, Tarski used to stress that any formalized theory consists of *meaningful* sentences. Let us quote a famous passage from the *Introduction to Logic*:

> "From time to time one finds statements which emphasize the formal character of mathematics in a paradoxical and exaggerated way; although fundamentally correct, these statements may become a source of obscurity and confusion. Thus one hears and even reads occasionally that no definite content may be ascribed to mathematical concepts; that in mathematics we do not really know what we are talking about, and that we are not interested in whether our assertions are true. One should approach such judgments rather critically. If, in the construction of a theory, one behaves as if one did not understand the meaning of the terms of this discipline, this is not at all the same as denying those terms any meaning. It is, admittedly, sometimes the case that we develop a deductive theory without ascribing a definite meaning to its primitive terms, thus dealing with the latter as with variables; in this case we say that we treat the theory as a FORMAL SYSTEM. But this situation (which was not taken into account in our general characterization of deductive theories given in Section 36) occurs only if it is possible to give several interpretations for the axiom system of this theory, that is, if there are several ways available of ascribing concrete meanings to the terms occurring in the theory, but if we do not desire to give preference in advance to any one of these ways. A formal system, on the other hand, for which we could not give a single interpretation, would presumably, be of interest to nobody."[63]

Such an explanation corresponds to the semantic shift in modern logic, which has been so much commented. Tarski did not initiate it from scratch,[64] but he turned

---

[61] See the rich materials recently published by P. Mancosu [46].

[62] Contribution to the discussion of P. Bernays, Colloque International de Logique, Bruxelles, 1953, *Revue Internationale de Philosophie* **27–28** (1954), 18–19; in Tarski [70, **IV**, p. 713] (my emphasis).

[63] Tarski [68, pp. 128–129].

[64] One source of the semantic shift is well identified by Wolenski's account: "Tarski grew up in a so-to-speak protosemantic atmosphere. The Lvov-Warsaw school was strongly influenced by the

it into a heuristic shift. Tarski wanted indeed the formalization be closely tied to concrete interpretations and not lead too far from "ordinary" or "normal"[65] mathematics, which used to make no reference to the syntax of the language.

c) Third, Tarski made a tight link between Hilbert's syntactic analysis of axiom systems and deductive proof on the one side and, on the other side, algebraic methods of logic as developed by Peirce, Schröder, Löwenheim and Skolem.[66] As Feferman wrote, Tarski "would axiomatize and algebraicize whenever he could".[67] In and of themselves, algebraic methods involve the correlation of a formal aspect induced by the use of variables and a semantic aspect anchored in the many interpretations we may possibly give to the variables. 'Meaning' is thus specified as 'interpretation', i.e. as 'model'. In Tarski's development of semantic methods converged the philosophical-logical semantic tradition, which originated from Brentano, and the trend of algebraic logic. This trend and its interpretative aspect were in fact present in Hilbert's *Foundations of Geometry* and in his development of metamathematics [54]. But what has been specific in Tarski's own contribution was the study of *the class of models* (all possible models) of a given formal system, instead of considering only one definite model.

d) Fourth, Tarski aimed at constructing a general theory of semantic concepts in a formal deductive way. For instance, he notably axiomatized the consequence operation. Now, what basic concepts his formal semantics consisted in? In "Grundlegung der wissenschaftlichen Semantik" [63], he wrote the following:

> "We shall understand by semantics[68] the totality of considerations concerning the concepts which, roughly speaking, express certain connections between the expressions of a language and the objects and state of affairs referred to by these expressions. As typical examples of semantic concepts we may mention the concepts of denotation, satisfaction, and definition [...] The concept of truth – and this is not commonly recognized – is to be included here, at least in his classical interpretation, according to which 'true' signifies the same as 'corresponding with reality'."[69]

Brentanist tradition ... [Brentano's] thesis that mental acts are intentional has in himself a semantic dimension. When Polish philosophers began to speak about names and sentences instead of presentations and judgments, this changed intentional relations into semantic ones, that is reference and truth. Moreover, the Brentano legacy decided that linguistic expressions were to be considered to be meaningful. This aspect of language almost automatically invited semantic studies." Wolenski [84, pp. 10–11]. Another well known source was the development of mathematics, since at least the emergence of non-Euclidean geometries (for more see Webb [78]).

65 Tarski [60], English translation in Tarski 1983, p. 111.

66 Tarski's main technique, the elimination of quantifiers, is an outstanding example of the confluence of an usual practice of the algebra of logic with Hilbert's formulation of the decision problem. See the Introduction and the Notes 4, 5, 11, 21 of Tarski [64].

67 Duren [19, p. 402].

68 See Wolenski's historical account: "The word 'semantic' became popular in philosophy in the thirties.. . .Poland was an exception in this respect. In *the twenties* Polish philosophers began to use the word 'semantyka' for considerations for the meaning-aspect of language." Wolenski [84, p. 1] (my emphasis).

69 Tarski [63], in Tarski [69, p. 401].

It is clear that Tarski aimed at building a theory of reference, and not at a theory of meaning. 'Meaning' is not a semantic term in Tarski's formal semantics.[70] We have to keep in mind this fundamental feature, in order to understand correctly some consequences we shall discuss later.

e) Fifth, in studying deductive theories from the semantic point of view, one has therefore to study *the semantics of formal systems*. This study constituted a new direction of metamathematics. It consisted of examining the interconnections between syntactic properties of formal systems and mathematical properties of their models. The type of problems Tarski considered was the following:

> "Knowing the formal structure of axiom systems, what can we say about the *mathematical* properties of the models of the systems; conversely, given a class of models having certain *mathematical* properties, what can we say about the formal structure of postulate systems by means of which we can define this class of models? As an example of results so far obtained I may mention a theorem of G. Birkhoff (*Proceedings of the Cambridge Philosophical Society* **31**, 1935, 433–454), in which he gives a full mathematical characterization of those classes of algebras which can be defined by systems of algebraic identities. An outstanding open problem is that of providing a mathematical characterization of those classes of models which can be defined by means of arbitrary postulate systems formulated within the first-order predicate calculus".[71]

As is well known, Tarski defined the concept of model [62, 63],[72] which was informally employed by many previous mathematicians and logicians. He paid attention to the relations of a language to its models and, inversely, of a class of models to a set of axioms able to express the formal theory of the class under consideration. This back-and-forth method between axioms systems and classes of models constituted Tarski's original way of practicing "conceptual analysis" for *mathematical purposes*, though it had been introduced and mainly used by modern logicians, notably by Frege, Russell, and Hilbert for *foundational* purposes. As a result of this new way of thinking, model theory came into being.

f) Sixth, as a further consequence of the semantic-heuristic shift he achieved, Tarski claimed that there was no universal formal language, no universal metatheory for the whole domain of mathematics. As early as 1930, he observed that "strictly speaking, metamathematics was not to be regarded as a single theory. For the purpose of investigating each deductive discipline a special metadiscipline should be constructed". This is contrary to the logicist view holding that logic is the universal metalanguage. Hilbert had assumed relativism *within* mathematics, since he

[70] According to Quine's later account "The main concepts in the theory of meaning, apart from meaning itself, are *synonymy* (sameness of meaning), *significance* (or possession of meaning) and *analyticity* (or truth in virtue of meaning). Another is *entailment*, or analyticity of the conditional. The main concepts in the theory of reference are *naming, truth, denotation* (or truth-of), and *extension*. Another is the notion of *values* of variables." *From a logical point of view*, Cambridge (Mass), 1953, p. 130 (quoted after [84]).

[71] Contribution to the discussion of P. Bernays; in Tarski [70, **IV**, p. 714] (my emphasis) – The open problem is the definition of elementary classes, the solution of which will be given later through the method of ultraproducts.

[72] For a recent historical account of this concept in Tarski's work see Mancosu [47].

stressed that a proof was relative to the chosen set of axioms for the theory under consideration.[73] But, on the logical level, not only had Hilbert never explicitly disclaimed the view of (syntactic) logic as being the universal language, but he also suggested his own conception of proof theory should succeed where Frege's failed, since it aimed at giving a consistency proof for a formal system of arithmetic. We know that Gödel's second incompleteness theorem (1931) destroyed this aim, at least in the form and scope Hilbert ascribed to it. Developing semantic considerations and stating the distinction language/metalanguage as the king road to avoid antinomies led Tarski to *a logical relativism*, namely a *semantic relativism*: semantic concepts "must always be related to a particular language".[74] However and at the same time, Tarski thought that the concepts of logic penetrate the whole domain of mathematics and that the methodology of deductive sciences is "a general science of sciences". Logic, wrote Tarski, is "a discipline which analyzes the meaning of the concepts shared by *all* the sciences, and states the general laws ruling those concepts".[75] It is clear that logic is here not only a very fruitful tool for getting new mathematical results, but the tool "par excellence" for laying the basic laws of general semantic concepts which are involved in the analysis of deductive theories. This sounds like a kind of logicism, namely a *semantic logicism*, in comparison with Frege and Russell's syntactic logicism. One may see a tension or even a conflict[76] between Tarski's semantic relativism and his semantic logicism. And a similar tension exists also in respect to other issues on which Tarski, nearly at the same time or even in the same paper,[77] sustained views seemingly not fully compatible with each other. For the point that we are now discussing, I think that the "tension-problem" has been resolved by Feferman's detailed analysis of the two sides of Tarski's efforts.[78] Feferman argued that Tarski was first and foremost a mathematician and that he actually took a straightforward, though first informal, *model-theoretic way* since at least 1924; therefore he used the notions of definability and truth in a relative sense, as he undoubtedly did in his paper on the definable sets of real numbers (1931). On the other hand, "Tarski thought that as **a side result** of his work on definability and truth in a structure, he had something important to tell the philosophers that would straighten them out about the troublesome semantic paradoxes such as the

[73] Hilbert [30, p. 169]: "the concept 'provable' is to be understood as relative to the underlying axioms system. This relativism is in accordance with the nature of things and necessary."
[74] Tarski [63], in Tarski [69, p. 402].
[75] Tarski [68, p. XII] (my emphasis, in order to stress that here Tarski meant not only the deductive sciences, but also the experimental sciences). The scope of logic is even wider, since Tarski aimed to create "a unified conceptual apparatus which would supply a common basis for the whole of human knowledge". See S. Feferman's comments in Feferman [23].
[76] J. Wolenski [83, p. 331].
[77] That happened at least twice: in Tarski [68] and in Tarski [66, Sections 22 and 23].
[78] Feferman [22].

Liar, by locating for them the source of those problems..." In the *Wahrheitsbegriff* (1933/36), according to Feferman, "we are not talking about *truth in a structure* but about *truth simpliciter*, as would be appropriate for a philosophical discussion, at least of the traditional kind". But the idea of a universal logical language is abandoned in the famous Postscript,[79] and, over time, Tarski qualified the logicist aspect of his first claims on the universality of logic. This is particularly clear in the way he answered the question 'What are logical notions?' [71], that we will discuss below (3.2 and Section 3.5). Moreover, Tarski always kept considering the whole domain of logic as a branch of "ordinary" mathematics and giving much evidence for his opinion through considerable work, even if he was willing to grant that the part of logic which is mathematics "does not *perhaps* exhaust logic".[80]

Another example of how Tarski moved far from the logicist stance is his treatment of type theory. As we know, Tarski used, in an informal way, the language of the simple type theory in his early essays, for instance in the paper on the definable sets of real numbers and in the *Wahrheitsbegriff*. That certainly represented an acknowledgement of Russell's logical program. But, it is well known too that Tarski preferred set theory, with just one type of individual variables, and came to abandon type theory in favor of the latter.[81] Therefore, he replaced *logical universality* with *mathematical universality*. It would be fine here to comment on F. Rodriguez-Consuegra's useful ramification of the concept of universality, which has been first suggested by Hintikka [39, pp. 13–15]. But, for my purposes, I need only to subscribe to the following point: on account of Tarski's footnote 2 to his *Wahrheitsbegriff*[82] and of his 1995 posthumous paper [73], F. Rodriguez-Consuegra argues that Tarski regarded more and more the language of set theory as a *mathematically universal* language with one *universal domain* of individuals.[83] It should just be added that Tarski regarded more and more the language of a sort of general algebra as fitting better his ambition to yield a universal language for mathematics, which would eliminate the current problems of set theory. He proposed already in 1953 a formalization of set theory without variables.[84]

---

[79] Feferman [22, p. 94]. While recognizing this fact, Feferman maintained for reasons that cannot be detailed here that, in the *Wahrheitsbegriff*, Tarski was after the concept of absolute truth (Feferman's emphasis, bold types are mine).

[80] Tarski [74, p. 27] (my emphasis).

[81] See Carnap's account in Mancosu [46, pp. 335–336]: "The Warsaw logicians, especially Lesniewski and Kotarbinski saw a system like PM – Principia Mathematica – (but with simple type theory) as the obvious system form. This restriction influenced strongly all the disciples; including Tarski until 'The Concept of Truth' (where the finiteness of the level is implicitly assumed and neither transfinite types nor systems without types are taken into consideration; they are discussed only in the Postscript added later). Then Tarski realized that in set theory one uses with great success a different system form. So he eventually came to see this type-free system form as more natural and more simpler".

[82] English translation, Tarski [69, p. 210].

[83] F. Rodriguez-Consuegra [53]. See also Feferman [22, 23], and Hintikka [39].

[84] Tarski [70, **IV**, pp. 605–606].

2.3 Tarski's Permanent Formal Leanings

The most striking trait of the formal way of working is certainly the search for invariant elements under changing conditions. This is a typical method in algebra. Tarski applied it in semantics as well.

Tarski had indeed a permanent attraction for purely algebraic methods and their potential links with logical operations. He invested much work in the rigorous algebraic reformulation and generalization of classical theorems, e.g. Sturm's theorem (on how many real roots a polynomial has in a given interval) that he transformed into a quantifier elimination principle.[85] – One has to point out, in passing, the finitistic character of this principle. – Tarski was also strongly interested in algebraic structures modeling logical operations, especially in Boolean algebras and cylindric algebras. He developed (together with Steven Givant) an algebraic approach to set theory which dispenses with variables: this general algebra was conceived of to provide a basic language for the whole field of mathematics. Algebra represented for Hilbert a paradigm for formal processing and extending the domain of surveyable objects. Tarski sought in it the means to avoid the logic of quantification. Hilbert introduced the transfinite axiom in order to justify the use of quantifiers, Tarski found out a mathematical device (Sturm's theorem) to eliminate quantifiers in the elementary theory of real numbers and Cartesian geometry.

3.1. A first example of Tarski's use of an invariant style is his semantic definition[86] of completeness: a theory is complete iff all its models are elementarily equivalent, i.e. iff a first-order sentence which is true in one model is also true in any other model of the theory. In other words, a theory is complete iff the set of first-order sentences that are proved in terms of one particular model remains invariant, so that one does not need to prove them again within another model. Tarski proved what was at his time an impressive result: the completeness of the first-order theory of real numbers and Cartesian plane geometry. As a consequence, he deduced that every first-order theorem about real numbers is already satisfied by algebraic real numbers. Thus, from a first-order logical point of view there is no difference between the field of algebraic real numbers, the underlying set of which is *countable*, and the field of real numbers, the underlying set of which is *uncountable*. This result may be considered as a corollary to Löwenheim's theorem that two non-isomorphic structures can be indistinguishable from the point of view of first-order logic (cardinality is neutralized). But Tarski shed a new light on it, presenting it as a logical invariance principle, which is weaker than algebraic isomorphism though none the shallower. Indeed, under the name 'transfer principle', it would have a great future and play a remarkable role not only in model theory, but also in some other mathematical branches: algebra, real algebraic geometry and analysis among others.

---

[85] Tarski [64, 65] (see [2, Part I, II and IV]).

[86] The syntactic definition is the following: a theory is complete iff every sentence of the language of the theory is provable or refutable. For first-order theories the two definitions are equivalent.

Early on, Tarski aimed at constructing a general theory of the equivalence relation involved in this principle. The notion of elementary equivalence appeared in print in the appendix to the second part of 'Grundzüge des Systemenkalküls' [61]. Tarski was then aware that he opened up a wide realm of investigation, and he proposed to carry out with mathematical methods. Ten years later, closing his Address at the Princeton University Bicentennial Conference, Tarski put forward the notion and suggested again further study of the subject. Later on, he gave an outline of the theory of elementary classes [67] and elaborated, in collaboration with R. Vaught, the notion of elementary extension (1957).

3.2. A second well known example is his explanation of the notion of logical operation in the type structure over a basic domain of individuals. This is to be found in the posthumous paper edited by J. Corcoran [71], in which Tarski addressed the following question: "What are Logical Notions?". Tarski's procedure was to extend to the domain of logic Felix Klein's *Erlanger Programm* (1872) for the classification of geometries according to their invariant elements under some group of transformations. For instance, the notions of metric Euclidean geometry are those invariant under isometric transformations, the notions of projective geometry under projective transformations, etc. Tarski proposed to consider logic as an invariant theory[87] and logical notions as those invariant in respect to any automorphism of the basic domain (any permutation of the domain) of the chosen universe of discourse. Considering a notion as logical depends on which formal language one chooses to define the term denoting this notion. Thus, if the formal language is that of type theory as developed by Whitehead and Russell in *Principia Mathematica*, then every notion is logical. Indeed, in this frame, set theory, within which the whole of mathematics can be constructed, is simply a part of logic, since the membership relation ($\in$) is invariant under the extension to higher types of any permutation of the domain of individuals. Thus, it appears that type theory was built in such a way as to justify logicist reductionism. Otherwise, if the language for formalizing set theory is Zermelo's first-order system – in which we have no hierarchy of types, but only one universe and the membership relation between individuals as *a primitive term* –, then mathematical relations are *not* logical. Indeed, the membership relation is *not* logical, since the only binary relations invariant under any permutation of the basic domain are the empty relation, the universal relation, the identity relation and its complement.[88] Tarski concluded his essay stressing that the given definition did

[87] Feferman [20, footnote 5], noted that Tarski seems to have been unaware of the first proposal of that type by F. I. Mautner, An extension of Klein's Erlanger Programm: logic as invariant theory, *American Journal of Mathematics* **68** (1946), 345–384. On his side, P. Mancosu states in his recent paper [46] that the idea of using Klein's strategy was first suggested by Alexander Wundheiler on the ground of a method expounded by Tarski and Lindenbaum, Über die Beschränkheit der Ausdrucksmittel deduktiver Theorien, *Erg. Math. Koll.*, **VII** (1936), 15–22. Wundheiler took part in the 10 January 1941 meeting, which was one of the series Tarski, Quine and Carnap had together during the academic year 1940–1941 at Harvard.

[88] V. McGee showed that the logical operations in Tarski's sense are exactly those which are definable in the language $L_{\infty,\infty}$: Logical operations, *Journal of Philosophical Logic* **25** (1996), 567–580, quoted after Feferman [20]. Solomon Ferferman bases on McGee's result two objections. The

not, in and of itself, imply a definite answer to the addressed question. Once again he emphasized that his logical work was free from any philosophical opinion, – which naturally does not mean free from set-theoretic methods. Conversely, technical results did not, by themselves, settle philosophical questions connected with them. That is to say that, in Tarski's view, the connection between logic and philosophy is a one-to-many relation. Tarski's emphatic and persistent professed neutralism towards philosophical views and his pluralism (that Wolenski called "liberalism") match this kind of connection and suggest a rather positivist philosophical attitude.

If a characteristic way of formal thinking is first and foremost reasoning in terms of variables (having many possible meanings) and invariants (under such or such transformation), then Tarski was a very enthusiastic "formalist" mathematician, in a sense, however, which encompasses none of the three main features that Brouwer highlighted in his 1912 essay (see above 1.2). Tarski indeed dealt with meaningful sentences, understood consistency in the sense of satisfiability by a model, and alleged that he would not support a Platonistic existence for abstract entities. This apparently paradoxical result has a twofold explanation: (1) Tarski really provided formalism with a new substance, (2) Brouwer's influence really contributed, even if by no direct and not always acknowledged ways, to important aspects of the shift from a relatively dominant syntactic view to the alliance of syntax and semantics.

3.3. Early on, Tarski asserted that the union of syntax and semantics, that he initiated, could be "theoretically" placed under the spirit of Lesniewski's "intuitionistic formalism",[89] while he claimed at the same time the independency of his technical achievements from any philosophical view. As it seems clear from the expression coined from the two previously contrasted terms, the "intuitionistic formalism", assumed as "an agreement in principle"[90] with Lesniewski's standpoint, might have been also a way to achieve the conciliation, initiated by Hilbert, between Brouwer's demand for contentual constructs and the formal processes of axiomatic and logic. That does not mean that Tarski accepted Brouwer's philosophical subjectivism, according to which one has to completely separate mathematics from language, especially from its description by logic, and to recognize mathematics as a "languageless activity of the mind having its origin in the perception of a *move of time*",[91] which constitutes the basic *Urintuition*. Moreover, working to bring closer logic and "ordinary" mathematics, Tarski could not share the idea of a separate autonomy for each of the two domains, upon which Brouwer insisted.

---

first one is that Tarski assimilates logic to set-theoretical mathematics, what was indeed Tarski's own permanent aim. Second, Tarski failed to explain logicality across domains of different sizes. Feferman proposes a homomorphism invariance criterion to correct this failure. He also mentions the proof-theoretic approach of J.I. Zucker and R.S. Tragesser (The adequacy problem for inferential logic, *Journal of Philosophical Logic* **7** (1978), 501–516), which leads to characterize logical operations as exactly those of the first-order predicate calculus.

[89] Tarski [59], in Tarski [69, p. 62].

[90] Tarski quoted the page 78 of S. Lesniewski, Grundzüge eines neuen Systems der Grundlagen der Mathematik, Sections 1–11, *Fundamenta mathematicae* **1** (1929), 1–81.

[91] Brouwer [13, p. 510].

To stress the contrast with Brouwer's *Urintuition*, J. Wolenski replaced 'intuitionistic formalism' by 'intuitive formalism' in his account of Lesniewski's systems.[92] Now, the word 'intuition' may rapidly induce a philosophical commitment, either to intuitionism – in a purely subjective option – *or* to Platonism if one holds that the subjective intuitive faculty is connected with an objective independent world *or also* to some kind of Kantian *a priori* as it was differently understood by mathematicians, for instance by Poincaré, Brouwer, Hilbert. But Tarski did not elaborate any specific theory of intuition. As Wolenski pointed out to me, Lesniewski and Tarski understood 'intuition' "quite customary, namely as an ability to grasp contents (meaning)". Therefore, it seems to me more appropriate to characterize Tarski's real way of working simply and merely as a *semantic formalism*, in opposition to the syntactic formalism shared by Frege, Dedekind and, to some extent, Hilbert. Those three hoped first and foremost to catch the *entire* content of a mathematical theory through a logical analysis of the syntactic properties of its fixed axioms system, while Tarski aimed at knowing under which logical conditions one can *extend the content* of a definite model of the theory.

Anyway, Tarski changed his mind: in a footnote added in 1956 in the English translation of his essay he pointed out that the "intuitionistic formalism" could no longer appropriately mirror his new attitude. What was no more convenient in this expression: 'intuitionistic', 'formalism' or both? Unfortunately, Tarski did not go so far as to positively describe what his new attitude was. Did Tarski keep silent because he separated philosophical thinking from scientific logical work? Certainly yes, even though there might have been other reasons.

3 Tarski's Philosophical Pluralism

Now, it becomes difficult to say that Tarski's *explicitly assumed* philosophical attitude matched the undoubtedly formal orientation of his practice. The path from the latter to the former is not straightforward. And that is not astonishing, since Tarski aimed to disconnect scientific reasoning from philosophical principles and, therefore, thought that a mathematical or logical technique made no philosophical point of view mandatory. Wolenski's judgment is right: it was *not* a problem for Tarski that his philosophical attitude did not fully agree with his own research practice in logic and mathematics.[93] On the one hand, I do confirm the agreement between what I have called 'semantic formalism' and Tarski's actual practice. 'Semantic formalism' seems to me the right expression to characterize how Tarski actually worked. But, on the other hand, we have to take into account the following facts: (1) Tarski changed

[92] Wolenski [82, p. 145]. According to Wolenski, Lesniewski was a "radical formalist in the sense of requiring an unambiguous codification of the language of a given formal system", but he firmly rejected the conception of logic and mathematics as a game of symbols devoid of meaning. More generally, the *interpretative style* of cultivating logic in the Warsaw School went back to Twardowski's tradition.

[93] Wolenski [82, p. 192] (my emphasis).

his mind and upheld, tacitly or explicitly, *different* philosophical attitudes without explaining the reasons of those changes. (2) Moreover, he used to propose on the same issue, *at the same time*, several options and to leave the choice open. This causes us a relative embarrassment. A way out is indeed to consider that Tarski was willing to construct arguments, not to give free rein to his belief. Therefore, he was trying different consistent arguments, as it was usual in the Ancient philosophical-logical tradition, at least in the part called 'dialectic' by Aristotle. Tarski's alleged philosophical neutrality was actually a real and very commendable philosophical option. In my view, it is perhaps the only tenable, though uncomfortable, option. After all, philosophical thinking is not just adapting argumentation to prior belief.

That being said, we still have the task to distinguish what Tarski claimed explicitly from what he did in fact, and to take into account the arguments he developed as dialectic exercices or, with a more modern scientific term, as '*Gedankenexperimente*'. *Grosso modo*, one might say that, while he kept an anti-metaphysical general attitude (inherited from the Lvov-Warsaw School and strengthened by contacts with the Vienna Circle), Tarski stood on at the junction point of at least three views: a self-evident, though non-explicitly advocated, semantic realism, a strong logical nominalism with finitistic requirements, that he supported but moderately practiced, and an effective pragmatism, which finally permeated different levels of his thought.

3.1 Tarski's Explicit Rejection of Ontological Realism

Tarski's well known definition of truth is the classical one: truth as "correspondence" with reality. But what sense has to be given to "reality"? Tarski (1933) [66] rejected the realistic interpretation of his definition, in particular Gonseth's reproach of uncritical realism, i.e. of pre-Kantian realism. Tarski argued that classical formulations of the adequacy-relation between truth and reality, which are assumed to convey a realistic conception of truth, are neither precise nor clear enough. He preferred Aristotle's formulation, that he carefully recalled.[94] And he recognized that his *formal* definition corresponded to the *intuitive content* of Aristotle's formulation. But he claimed that there was no necessary bound between his semantic definition and any of the following standpoints: realism, idealism, empiricism, metaphysical attitude. That means that Tarski did not *base* his semantic explanation on *a priori* or initial realistic (nor idealistic, nor empiricist nor metaphysical) assumptions; – if one seeks philosophical understanding, it would perhaps be better to go the other way around: to get a philosophical understanding, and probably not a one-sided one, from the scientific explanation. Tarski's explanation does not give a criterion to confront the sentence 'snow is white' to the real factual conditions under which we may affirm or not the sentence under consideration. The explanation shows the *equivalence*

[94] "To say of what is that it is not, or of what is not that it is, is false, while to say of what it is that it is, or of what it is not that it is not, is true". (Tarski [62], in Tarski [69, p. 155, footnote 2]). It is worth noting that Tarski pointed out (p. 153) that he set aside, for instance, the utilitarian conception, according to which 'true' reduces to 'useful'.

between two sentences, traditionally referred to as the T-schema: the sentence 'snow is white' is true iff snow is white. Right to 'iff' we have a sentence and left to 'iff' we have the same sentence between quotation marks, i.e. we have the name of the sentence. *We do not go out of the universe of discourse*; we stay on a purely semantic level. The semantic definition states what truth is, not how to confirm or infirm it. Tarski meant that such a formal and non-effective definition needed not be backed up by a metaphysical or an epistemological conception. Semantics is indeed a scientific theory in its own right, and as a scientific theory it is supposedly philosophically neutral. Even if, with some right, one takes Tarski's claim concerning the neutrality of semantics *cum grano salis*, it should be taken for granted that Tarski rejected that pre-Kantian form of philosophical realism, which is also named 'essentialism' or 'ontological realism'.

3.2 Tarski's Possible Acceptance of "a Moderate Platonism" and Actual Semantic Realism

Tarski wrote indeed that he was never able to understand what is "the essence" of a concept.[95] This means that a definition of a concept does not aim to capture, in a Platonist style, the essence of what is designated by the concept. Indeed, when Tarski set a definition for a notion (truth, logicality), he constantly insisted upon the fact that his definition, constructed within the frame of theoretic semantics, suited the effective meaning or use of the notion. Clearly enough this indicates that Tarski deliberately kept distance from Plato's way of constructing "essential" definitions.

But, Platonism does not only consist in the search for "essential definitions". It means also the belief in an ideal existence of the essences assumed to be the objects of such definitions, namely the belief in the autonomous existence of abstract entities.

Now, it is not a paradox to claim that a formal way of doing mathematics and logic may lead to some form of Platonism. We have seen above several degrees in the scale of Plato's assumptions analyzed by Bernays, the top of the scale being reached by set theory. For his part and on the one hand, Tarski *used* abstract methods and set-theoretic concepts involving infinitistic and non effective ways of reasoning. This might have implied a positive *affirmation* of the ideal existence of those abstract entities. But Tarski never committed himself to such an ontological statement. He did not admit the usual Platonistic understanding of the axioms of set theory, according to which sets exist independently of any human constructions. Moreover, he generally did not use predicate variables or higher types in his metamathematical analysis of mathematical theories, and he restricted himself to first-order language, in accordance with his algebraic bent, which led him to his quantifier elimination

[95] Tarski [66, Section 18].

technique.[96] An interesting interpretation sees in Tarski's attitude an "as-if-realism", that is to say that Tarski mathematically behaved *as if* abstract entities existed, though his philosophical stand imposed restriction to individuals.[97] This interpretation may have some loose connection with Tarski's acknowledgement in the discussion period for a 1965 meeting on the philosophical significance of Gödel's incompleteness theorems.[98] Tarski said indeed that he, "perhaps in a 'future incarnation', would be able to accept a sort of moderate Platonism". In all likeliness Tarski said that he would accept a milder version of Gödel standpoint, which was an outright Platonism. This version could consist, for instance, in accepting *only* the sequence or the totality of natural numbers.[99] Furthermore, Tarski meant he would "accept", not advocate.

On the other hand, the search for invariant principles might include the philosophical question about *identity* and *persistence* of some mathematical or logical content. For instance Tarski's transfer principle allows to transpose one and the same content from one model to another irrespective of the particular formal setting in which it is encapsulated. The transfer principle points out the persistency of a property, a meaning, through different formal frames. It is not a formal extension principle from intuitive operations to abstract ones ("formal permanency law" in the language of the XIXth century), but an extension of the same concrete meaning to

---

[96] I did stress [2, Parts II and IV] the link between the idea of quantifier elimination and Hilbert's achievements, both on geometry where the aim was to determine the scope of the continuity axioms, the independency of which he proved through the construction of a non-Archimedean model, and on metamathematics, the goal of which was to check the consistency of formulas (ideal propositions) through the reduction of proofs to numerical equations or non-equations (real contentual propositions without variables).

[97] See Rodriguez-Consuegra [53, p. 240]. The schema of an as-if attitude is already present in Hilbert [31, p. 187]: "In my proof theory it is not asserted that one can always effectively pick up an object among infinitely many objects, but that one can always, without risk of mistake, do as if the choice were made" (p. 187). See also Bernays [6], in Bernays [8, p. 60], in Mancosu [44, p. 262]: "The view at which we have arrived concerning the theory of the infinite can be seen as a kind of philosophy of the 'as if'. However, it differs entirely from the so-called philosophy of Vaihinger in the fact that it emphasizes the consistency and the stability [*Beständigkeit*] of the idea-formations...″. Mancosu [45, p. 316], noted that the same idea was previously developed by H. Behmann in his 1918 Dissertation (Hilbert was supervisor). The idea is still attractive for formalists. See Robinson [51], *Selected Papers*, **II**, p. 507: "My position concerning the foundations of Mathematics is based on the following two main points or principles. (i) Infinite totalities do not exist in any sense of the word (i.e. either really or ideally). More precisely, any mention, or purported mention, of infinite totalities is, literally, *meaningless*. (ii) Neverthelesss, we should continue the business of Mathematics 'as usual', i.e. we should act *as if* infinite totalities really existed" (Robinsons's emphasis). See also Robinson [52].

[98] Typescript of extemporaneous remarks during the discussion period for a symposium held in Chicago at a joint meeting of the Association of Symbolic Logic and the American Philosophical Association, 29–30 April 1965, Bancroft Library. Briefly quoted in Wolenski [83, p. 336]. A longer excerpt is quoted in Feferman [21, p. 61], and in Anita Burdman Feferman and Solomon Feferman [24, p. 52]. The topic is discussed at length in Rodriguez-Consuegra [53].

[99] This interpretation matches the requirement Tarski imposed on the construction of a nominalistic language. See Mancosu [46, p. 336], quoted below.

other formal languages. In *The completeness of elementary algebra and geometry* [64], Tarski noted that, in order to determine whether or not a classical theorem of geometry belongs to his elementary formal system, "it is only the *nature* of the concepts, not the character of the means of proof that matters".[100] What Tarski highlighted here is that an elementary (first-order) theory may encompass concepts expressible or provable under non-elementary conditions, which are known to be satisfied in some particular model of the (complete) theory, for instance in real numbers. That is to say, a first-order theory may capture much more properties than *first-order definable* properties. From a logical (technical) point of view, this fact is, in and of itself, significant. From an epistemological point of view, this fact means that understanding a concept is not reducible to the technique of reasoning about it in some well-defined frame. Last but not least, from an ontological point of view, the insistence on a mathematical content independent of its formal definability or its proof has undoubtedly a Platonistic flavor, even if we cautiously distinguish 'nature' from 'essence'. But how "the nature" of a concept has to be understood? The reasonable answer in the frame of Tarski's mode of work seems to me the following: just as truth is not exhausted by deductive verification (Gödel's incompleteness theorem and Tarski's undefinability theorem), meaning is not exhausted by formal expression.

But again what is 'meaning'? This is a philosophical issue, which Tarski did *not* tackle. As we saw above, formal semantics did not comprise a theory of meaning. Wolenski pointed out that Tarski did *once* in 1936 made a remark on the subject in a discussion of a paper by M. Kokoszinska.[101] Tarski simply observed that the concept of formal language was clearer and logically less complicated than the concept of meaning. But Tarski *showed* (notably trough the transfer principle) that meaning transcends formal language. This naturally leads to a realist view of meaning, in the same sense as the undefinability theorem leads to a realistic understanding of truth, *in contrast with a constructive view*. Now, as noted above, Tarski did reject the possibility of a logical link between semantic results and philosophical assumptions. He did reject metaphysical realism. Is there some real tension or, as Wolenski wrote some "cognitive dissonance"?[102] I do not think so, at least concerning this specific

---

[100] Tarski [70, **IV**, pp. 305–306] (my emphasis). The remark is repeated in the later in California published version [65], Tarski [70, **III**, p. 307].

[101] I thank Professor J. Wolenski for drawing my attention to this remark and for the translation of it from Polish; see Tarski [70, **IV**, p. 701]. The discussion took place in Krakow during the 3rd Polish Philosophical Congress after Kokoszynska's talk 'Concerning relativity and absoluteness of truth', and the translation of the remark is the following:

"It follows from the words of the speaker [that is, M. Kokoszynska – J. W.] among others things that the concept of truth – in one of its interpretations – should be relativized to the concept of meaning. Would not be simpler to relativize to the concept of language, which is clearer and logically less complicated than the concept of meaning?

Kokoszynska replied that the concept of language implicitly involves the concept of meaning. Hence, a double relativization should be made (1) to the stock of shapes or sounds; (2) to meaning".

[102] Wolenski [82, p. 192].

point. If we adopt Wolenski's distinctions between different kinds of realism, in particular between metaphysical realism and semantic realism,[103] we may say that Tarski's views on truth and on mathematical concepts pertain to semantic realism, not to ontological realism. Moreover, as Wolenski showed, Tarski's semantic realism *does not imply* metaphysical realism, just as Tarski himself claimed. This explains why Tarski could uphold at the same time a realist attitude within the semantic sphere and a dislike of Platonism, which is, he thought, "unsatisfactory as an end-point in philosophical analysis".[104]

3.3 Logical Nominalism

In fact, Tarski invested a valuable amount of energy to avoid Platonism. – As for intuitionism or logicist reductionism he apparently felt no need to keep clear of them. – Influenced by Lesniewski and Kotarbinski, Tarski developed a strong nominalistic bent. It is worth recalling that nominalism emerged in the Middle Ages in the debate about universals and particulars. According to Mycielski [48], Tarski was familiar with this debate through Twardowski's book *Six Lectures on Medieval Philosophy* and with the distinctions between nominalism, Platonism and conceptualism. To clarify things, I recall a brief characterization. Nominalists admitted only the existence of particulars. Conceptualists admitted the existence of concepts or forms, especially when the universals were represented in individuals. Platonists admitted the existence of concepts and forms independent of human mind. What distinguished conceptualists from nominalists is that they did not reduce concepts to mere signs or names: concepts were contentual operations of thought; then, their existence was understood as a thought-existence. What distinguished conceptualists from Platonists is that they did not detached the existence of concepts from the operating thought: concepts did not exist on their own, prior to thought, they did not play the role of the essences of empirical things.[105] In the view of this tripartition, it seems to me that one could consider conceptualism as very near to semantic realism, despite the fact that Tarski spoke neither of conceptualism nor of semantic realism. His concern was to stress his opposition to Platonism. Indeed, Tarski described himself as a nominalist. In the typescript of the remarks at the 1965 symposium on Gödel's incompleteness theorems, Tarski said:

---

[103] Wolenski [85, pp. 135–148]. Wolenski defines semantic realism by the fact or supposition that meaning transcends use. Here I assume that semantic realism means also that meaning transcends language.
[104] Quoted by Mancosu [46, pp. 334, 348].
[105] We may note, in passing, that Hilbert's and Bernays' designation of Platonism as "conceptual realism" was literally adequate.

> "I happen to be, you know, a much more extreme anti-Platonist.... I represent this very crude, naïve kind of anti-Platonism,[106] one thing which I could describe as materialism, or nominalism with some materialistic taint, and it is very difficult for a man to live his whole life with this philosophical attitude, especially if he is a mathematician, especially if for some reason he has a hobby which is called set theory..."

Thus, Tarski avowed himself the tension between his philosophical views and his mathematical needs. And he maintained this duality (which supports the "as-if-Platonism" interpretation). Later on indeed, at the closing of his seventieth birthday symposium (1976), Tarski said:

> "I am a nominalist. This is a very deep conviction of mine. It is so deep, indeed, that even after my third reincarnation, I will still be a nominalist. ... People have asked me, 'how can you, a nominalist, do work in set theory and logic, which are theories you do not believe in?'... I believe that there is value even in fairy tales and the study of fairy tales".[107]

This might be interpreted as a joke. But a joke is also an usual way *not* to give one's last word on some issue.

In fact, there is a deep connection between Tarski's professed nominalism and his actual formal practice, which was strongly impregnated with an algebraic spirit. Tarski would have probably not disapproved Brouwer's judgment, according to which abstract entities exist only "on paper". Working in set theory does not necessarily mean believing in a hypostastic existence of sets. After all, it is possible to deal with concepts without reifying them, i.e. without transforming them into "real" objects, "real" being interpreted either as having a material character through concatenations of signs or as a Platonist universe of timeless objects. But, was it not Tarski's aim to disconnect the semantic sphere, to which mathematical concepts belong, from the ontological one and, therefore, at eliminating unnecessary ontological suppositions? Naturally 'yes', and we have even stressed that semantic realism does not necessarily entails ontological realism.

Nevertheless, we need, I think, to have an idea of the philosophical status that Tarski might have attributed to meaning, which he used as an informal notion. From the *model-theoretic point of view*, 'meaning' is 'interpretation' or 'realization' (and truth is equivalent to the existence of a model). Now, there are interpretations with infinite basic domains. Then, in Tarski's mind, what would have been the satisfactory *philosophical* final view on interpretations/meanings of the abstract theories, i.e. his final view on abstract *entities*? Might Tarski have accepted, in accordance with the formalist tradition, abstract entities as beautiful and fruitful fictions ("fairy tales"), something similar to Leibniz differential operator or to Hilbert's ideal elements, the justification of which is the ultimate reduction to finite entities? If the answer were 'yes', then Tarski's position would result in a combination of

---

[106] See also Mycielski [48, p. 217]: in 1970 Tarski mentioned to Mycielski "the Platonic belief of Gödel that sets can be seen (seen, not imagined) in our minds almost like physical objects", and added that this belief "is bewildering".

[107] Anita Burdman Feferman and Solomon Feferman [24, p. 52]. Also Mycielski [48, p. 216]: "Tarski told me that he is a nominalist".

nominalism and finitism. As we shall see in the next paragraph, some evidence is now available for associating Tarski's nominalism with finitism.

But, from the *philosophical point of view*, has Tarski really thought that meaning belongs to the world of fairy tales? Was meaning, in Quine's words, a myth? Would Tarski have agreed with Quine's reductionism, and would he have *ultimately* admitted an elimination of meaning in favor of its linguistic medium, that he found clearer? I do not think so, because accepting the linguistic reduction of meaning would tip the whole enterprise of formal semantics into a mere linguistic analysis, what it *is not*. Then, might Tarski have considered meaning as a mental act or process? A positive answer to this question would lead him near either to the medieval conceptualism or to modern intuitionism. But, on the basis of the available evidence relative to his cultural background, we cannot suppose that Tarski would have accepted to go Brouwer's road. On the other hand, Tarski did not express himself about conceptualism. Then the question of what acceptable philosophical status could be given to meaning from Tarski's point of view remains open.

Now, how can we understand Tarski's alliance of nominalism with materialism in his claim at the Chicago meeting? On Wolenski's account,[108] nominalism *and* materialism (physicalism) were typical of Kotarbinski's reism. Wolenski thinks that Tarski was much more attracted to reism than Mostowski admitted[109] and he suggests understanding materialism as being an empiricism. Tarski stressed indeed that between logical and empirical statements "there is only a mere gradual and subjective distinction"[110] and that logical sentences might be just as revisable as the factual ones.[111] Thus, we have necessarily to take into account a "time coefficient" and to refer any hypothesis to a given historical stage of the development of a science. This empiricist basic option sheds substantial light on Tarski's nominalism and make a bridge with the logical empiricism of the Vienna Circle, but does not answer the question why fairy tales keep being attractive. In other words, how is it possible to reconcile semantic realism with logical nominalism?

3.4 Nominalism, Finitism, Constructivism

Linked with his professed nominalism, Tarski upheld two other views, as we newly became aware through P. Mancosu's work on an important set of notes found in Carnap's Nachlass in Pittsburgh, the edition of which is being prepared by Greg Frost-Arnold. Carnap reported indeed that, during the Fall of 1940, he regularly met Quine and Tarski at Harvard, and discussed with them on the construction of a finitistic mathematical language for science. This language was intended to be

[108] Wolenski [82, Chapter XI].
[109] A. Mostowski, Tarski, Alfred, *The Encyclopaedia of Philosophy*, **8**, P. Edwards ed., 1967, New York, Macmillan, 77–81; Wolenski [83, footnote 2, p. 340].
[110] Quoted in Mancosu [46, p. 328].
[111] Tarski [72].

type-free: P. Mancosu highlights the shift that was taking place in Tarski's thought (and in logic in general) from type-theoretic to first-order languages.

In developing their project Carnap, Quine and Tarski agreed on three points: the language should be nominalistic, (weakly) finitistic and constructivistic. It is worth quoting after Mancosu the whole passage [46, p. 336]:

> "... We agreed that the language must be nominalistic, i.e., its terms must not refer to abstract entities but only to observables objects or events. Nevertheless, we wanted this language to contain at least an elementary form of arithmetic. To reconcile arithmetic with the nominalistic requirement, we considered among others the method of representing the natural numbers by the observable objects themselves which were supposed to be ordered in a sequence; thus no abstract entities would be involved. We further agreed that for the basic language the requirements of finitism and constructivism should be fulfilled in some sense. Quine preferred a very strict form; the number of objects was assumed to be finite and consequently the numbers occurring in arithmetic could not exceed a certain maximum number. Tarski and I preferred a weaker form of finitism, which left open whether the number of all objects is finite or infinite. Tarski contributed important ideas on the possible forms of finitistic arithmetic."

First of all, one notes that here 'nominalism' is understood in its medieval sense: only particulars were admitted. No mention was made of the modern sense given to the term by members of the Vienna Circle, especially Carnap, who advocated the view that mathematics is reducible to some syntax of language. I guess Tarski would not have supported this view. Nevertheless, the material published by P. Mancosu shows the driving role Tarski played in these discussions and the influence he had in the early development of twentieth century analytic philosophy.

Second, as Mancosu stresses, no clear distinction was made in the Carnap's notes between nominalism and finitism. On 10 January 1941, Tarski unfolded his view on finitism[112] by stating that he basically "understood" only languages, which satisfy the following conditions: finite (later on, he also allowed for infinite) number of individuals, the individuals are physical things (Kotarbinski's reism), there are no variables for universals (classes and so on), i.e. there is no Platonic assumption. Tarski brought a precision, which seems to me important, because it makes very clear how pivotal were his algebraic leanings. He added indeed: "Other languages I 'understand' only the way I 'understand' [classical] mathematics, namely as a calculus". This is an *explicit* acknowledgement of one of the basic views Tarski had from the beginning of his work: even if it was not until the 1950s that model theory flourished as a discipline in its own right, the model-theoretic view of mathematical language as an reinterpretable calculus has been permanently present in Tarski's mind and practice from the beginnings of his work. This algebraic view did not totally preclude the opposite view of set-theoretic language as a universal mathematical language. But it became more and more prominent, so that it led to the project of a general algebra as fundamental base for the whole mathematics. This

[112] Mancosu [46, p. 343].

project has been embodied in his posthumous book (together with Steven Givant): *A Formalization of Set Theory without Variables*.[113]

Now, one may wonder whether "the method of representing the natural numbers by the observable objects themselves which were supposed to be ordered in a sequence" really dispenses with *the set* of natural numbers, which is involved, at least potentially, in the notion of sequence. But for Tarski the distinction between potential and actual infinity was not an essential one.[114] The main problem for him was whether logic and mathematics, which are "an indispensable tool for scientific research in empirical science"..."can be constructed or interpreted nominalistically".[115] Since he wanted to have elementary arithmetic, Tarski suggested to reformulate Dedekind-Peano's axioms so that no axiom of infinity is included and to construct a recursive arithmetic.[116] He also chose a constructive definition of elementary arithmetic.

3.5 Effective Pragmatism or the Final View on Meaning

In his early period, Tarski sometimes and somehow defended the intrinsic interest of metamathematical research. For instance, he declared the following:

> "Being a mathematician (as well as a logician, and perhaps a philosopher of a sort), I have had the opportunity to attend many discussions between specialists in mathematics...I do not wish to deny that the value of a man's work may be increased by its implications for the research of others and for practice. But I believe, nevertheless, that it is inimical to the progress of science to measure the importance of any research exclusively or chiefly in terms of its usefulness and applicability. We know from the history of science that many important results or discoveries have had to wait centuries before they were applied in any field. And, in my opinion, there are also other important factors which cannot be disregarded in determining the value of a scientific work. It seems to me that there is a specific domain of very profound and strong human needs related to scientific research, which are similar in many ways to aesthetic and perhaps religious needs."[117]

But, at the same time, Tarski repeatedly stressed the independence of his technical results from any philosophical assumption and their mathematical usefulness. It seems to me that over time Tarski came closer and closer to the outlook most fitting the scientific practice in general, namely a pragmatist outlook. By pragmatism I understand here simply an attitude primarily determined by the ways and needs of actual mathematical practice. Pragmatism rests upon the primacy given to use, but does not necessarily entails utilitarianism, which says that 'true' is nothing more than 'useful'.

[113] Tarski and Givant [75].
[114] Mancosu [46, p. 345].
[115] Letter to Woodger, 21 November 1948, quoted by Mancosu [46, p. 347].
[116] For more see Mancosu, pp. 350–354.
[117] Tarski [66], Tarski [70, **II**, p. 693].

Indeed, from the 1950s onward, much as an "ordinary" mathematician, he raised the question of applicability of metamathematical methods in a very straightforward manner. In particular he strove to show that his theory of elementary classes "had good chances to pass the test of applicability ... [and to] be of general interest to mathematicians".[118] On many other occasions, Tarski professed taking the practice into consideration, especially when he aimed to set a precise definition for a notion, the meaning of which has been previously vague or understood only in an informal way. One of the constraints he placed upon the definition is that it has to match the mathematical or logical *use*. Defining itself may be just setting criteria for using the notion. Thus, in the above quoted lecture 'What are logical notions?', Tarski explained that answers to questions such as the one he addressed may be of different kinds. In some cases, one may give an account of the prevailing usage of the expression denoting the *definiendum*: this is a descriptive definition. In other cases, one may set criteria for future usage, relatively independent from the current usage: this is proposing a normative definition. Tarski claimed to have set in his paper a normative definition, namely to have suggested a possible use for the expression 'logical notion'. This possible use fits the mathematical use, which originates from Klein's outstanding procedure to distinguish various systems of geometry. Anyway, be it actual or potential, usage keeps to be one of the basic conditions that the construct of a definition must satisfy. Moreover, Tarski explicitly added that the aim of "catching the proper, true meaning of a notion, something independent of actual usage, and independent of any normative proposals, something like the platonic idea behind the notion" constituted, to his eyes, "so foreign and strange an approach", so that he would simply ignore it. Now, may one not infer from this passage and from some other brief remarks including those on the concept of definable sets of real numbers [60],[119] on the semantic definition of truth [62, 66] that I have quoted above, and on the characterization of semantic concepts[120] that, in Tarski's *philosophical final view*, meaning was use?

Whatever the answer to this question might be and so surprising the union of pragmatism and semantic realism might seem, the gradually more salient role of usage in Tarski's thought and practice, as well as his basic and permanent motivation to making logic useful for the working mathematician, allows one to claim that Tarski's *effective* philosophical attitude was in keeping with a kind of pragmatism. All fruitful methods are welcome, he thought and wrote. The study of fairy tales is worthwhile, because they can be submitted to experiments so that they gain a firm

---

[118] Tarski [67], Tarski [70, **III**, p. 473].

[119] English translation, Tarski [69, p. 112]: "We then seek to construct a definition. . .which, while satisfying the requirements of methodological rigour, will also render adequately and precisely the actual meaning of the term ['definable set of real numbers']".

[120] Tarski [63], in Tarski [69, p. 402]: "the task of laying the foundations of a scientific semantics, i.e. of characterizing precisely the semantical concepts and of setting up a logically unobjectionable and materially adequate way of *using* these concepts, presents no further insuperable difficulties [as soon as we take into account the relative character of these concepts]" (my emphasis).

ground in our culture and they manifestly are "very useful and very helpful in the development, in the progress achieved" [by mathematics, therefore by physics and other sciences].[121] They provide with important results, either theoretic ones, which permit a better intrinsic understanding of the subject under consideration, or technical ones which can be applied, through physics, to the external world. As a helpful means of investigation, fairy tales do not contravene empiricism and, precisely because we are aware that they lack reality, they are compatible with nominalism. Last but not least, fairy tales satisfy inescapable human needs.[122]

Conclusion

I used in this paper expressions such as 'semantic formalism', 'semantic relativism', 'semantic logicism', 'semantic realism'. Those expressions, which may seem at first sight either surprising or finally trivial, must not be taken as a mere trick. Actually, they are stressing again and again that Tarski's fundamental aim was to establish formal semantics as a new branch of metamathematics. As a consequence of his aim, Tarski was constantly highlighting the semantic aspect of any method he adopted and any view he defended, and he was also constantly concerned with establishing the scientific autonomy of formal semantics. He contributed mostly to develop by rigorous means and to let largely known the interpretative style of the Polish School of logic.

Thus, while developing formal methods in this interpretative style, Tarski was greatly concerned with the idea of keeping close to mathematical practice and of holding non-dogmatic philosophical views. He was willing to experiment different, and even opposite, ways of constructing mathematical and logical theories. According to Steven Givant, Tarski very early developed an experimental style of working. In particular, the seminar on mathematical logic conducted by Lukasiewicz, to which he participated in the years 1920–1924, was viewed as "a kind of logico-mathematical laboratory where [the participants] could conduct experiments in assessing the expressive and deductive powers of various theories".[123] Much later, Tarski claimed to be "quite interested in attempts at constructing set theory on the basis of some non-classical logics, *simply as an experiment*. We shall see to what it will lead".[124] "Try and see" seems to have been a guiding principle of his logico-mathematical experimentation, and it was thus natural to make many different attempts with no *a priori* expectation of the result. In a fundamentally empiricist and pragmatic way, Tarski managed to blend nominalism, which is the philosophical counterpart of a finitistic requirement, which in its turn matches his empiricistic or

121 Quoted by Rodriguez-Consuegra [53, p. 248].
122 Compare with Weyl 1925–1927, in Mancosu [44, p. 141]: "there is a theoretical need, simply incomprehensible from the merely phenomenal point of view, with a creative urge directed upon the symbolic representation of the transcendent, which demands to be satisfied".
123 Givant [25, p. 52].
124 Typescript of Tarski's contribution at the 1965 Chicago meeting. Quoted by F. Rodriguez-Consuegra [53, p. 250].

physicalistic fundamental perspective, with a semantic realism, which is needed not only to develop beautiful theories, but also to support the *semantic* view that truth is not just proof, and meaning not just language. If one stands on this view *at a philosophical level*, then one has to pay the price for it, and the least one is just *not to accept* the reduction of truth or meaning to something else, whatever it might be. But, if, practically, i.e. for the working mathematician, showing the truth is *nothing but* proving some assertion and if meaning is *only* use, in accordance with rules (already established or to be formulated), then pragmatic considerations become primary, even in the study of the world of fairy tales.

Acknowledgement I am grateful to Jan Wolenski and to Paolo Mancosu for useful comments on a previous draft of this paper and for drawing my attention to some references. I thank also Sten Lindström and the referee for many improvements.

References

1. Benacerraf P. and Putnam H. 1983, *Philosophy of Mathematics. Selected Essays*, Prentice Hall, Inc. Englewood Cliffs, New Jersey (first ed. 1964).
2. Benis Sinaceur Hourya. 1991, *Corps et Modèles, Essai sur l'histoire de l'algèbre réelle*, Paris, Vrin (second ed. 1999).
3. Benis Sinaceur Hourya. 1993, Du formalisme à la constructivité: le finitisme, *Revue Internationale de Philosophie* **47**, n 186, 4/1993: 251–284.
4. Benis Sinaceur Hourya. 2001, Alfred Tarski: Semantic Shift, Heuristic Shift in Metamathematics, *Synthese* **126**: 49–65.
5. Bernays P. 1922, Über Hilberts Gedanken zur Grundlegung der Arithmetik, *Jahresbericht der Deutschen Mathematiker-Vereinigung* **31**: 10–19. English translation in Mancosu 1998, 215–222.
6. Bernays P. 1930, Die Philosophie der Mathematik und die Hilbertsche Beweistheorie, *Blätter für Deutsche Philosophie* **4**: 326–367. Reprinted in P. Bernays 1976, 17–61. English translation in Mancosu 1998, 234–265.
7. Bernays P. 1935, Sur le platonisme en mathématique, *L'enseignement mathématique*, XXIV, n^{os} 1–2: 52–69. Reprinted in P. Bernays, *Philosophie des mathématiques*, Paris, Vrin, 2003, 83–104. English translation in Benacerraf P. and Putnam H. [1], 274–286.
8. Bernays P. 1976, *Abhandlungen zur Philosophie der Mathematik*, Darmstast, Wissenchaftliche Buchgesellschaft. French translation, Paris, Vrin, 2003.
9. Brouwer L.E.J. 1912, Intuitionism and Formalism, *Bulletin of the American Math. Society* **20**: 81–96. Reprinted in Brouwer 1975, 123–137.
10. Brouwer L.E.J. 1928, Intuitionistische Betrachtungen über den Formalismus, Koninklijke *Akademie van wetenschappen te Amsterdam, Proceedings of the Section of Sciences* **31**: 374–379. Reprinted in Brouwer 1975, 409–414. English translation of §1 in van Heijenoort 1967, 490–492.
11. Brouwer L.E.J. 1929, Mathematik, Wissenschaft, und Sprache, *Monatshefte für Mathematik* **36**, 153–164. Reprinted in Brouwer 1975, 417–428. English translation in Mancosu 1998, 45–53.
12. Brouwer L.E.J. 1948, Consciousness, Philosophy, and Mathematics, *Proceedings of the Tenth International Congress of Philosophy*, Amsterdam, 1948. Reprinted in Brouwer 1975, 480–494.
13. Brouwer L.E.J. 1952, Historical Background, Principles and Methods of Intuitionism, *South Africa Journal of Science*, Cape Town, July 1952. Reprinted in Brouwer 1975, 508–515.
14. Brouwer L.E.J. 1975, *Collected Works* **I**, A. Heytind (ed.), North-Holland.

15. Cantor G. 1966, *Abhandlungen mathematischen und philosophischen Inhalts*, Hildesheim, Georg Olms Verlagsbuchhandlung, 1966.
16. Dedekind R. 1888, *Was sind und was sollen die Zahlen?*, Braunschweig, Vieweg.
17. Dedekind R. 1890, Letter to Keferstein, 27 February 1890, English translation in Jean van Heijenoort (ed.), 1967, 98–103.
18. Dauben J.W. 1995, *Abraham Robinson*, Princeton, Princeton University Press.
19. Duren P. 1989, *A Century of Mathematics in America*, Part III, American Mathematical Society, Providence R.I., 393–403.
20. Feferman S. 1999a, Logic, Logics, and Logicism, *Notre Dame Journal of Formal Logic* **40**: 31–54.
21. Feferman S. 1999b, Tarski and Gödel: Between the Lines, in Tarski J. and Köhler E. (eds.), *Alfred Tarski and the Vienna Circle. Austro-Polish Connections in Logical Empirism*, Dordrecht/Boston/London, Kluwer Academic Publishers, 53–64.
22. Feferman S. 2003, Tarski's Conceptual Analysis of Semantical Notions, in A. Benmakhlouf (ed.), Sémantique et épistémologie, Casablanca, Editions Le Fennec, 2003, 79–108.
23. Feferman S. 2004, Tarski's conception of logic, *Annals of Pure and Applied Logic* **126**: 5–13.
24. Feferman A. and Feferman S. 2004, *Alfred Tarski: Life and Logic*, Cambridge, Cambridge University Press.
25. Givant S. 1999, Unifying Threads in Alfred Tarski's Work, *The Mathematical Intelligencer* **21**(1), 47–58.
26. Heyting A. 1930, Die formalen Regeln der intuitionistische Logik, Sitzungsberichte der Preussischen Akademie der Wissenschaften, Phys.-Math. Kl. 42–56. English translation in Mancosu 1998, 311–327.
27. Hilbert D. 1899, *Grundlagen der Geometrie*, Zehnte Auflage, Stuttgart, B. G. Teubner, 1968.
28. Hilbert D. 1905, Über die Grundlagen der Logik und der Arithmetik, *Verhandlungen des dritten Mathematiker-Kongresses* (Heidelberg, August 1904), Leipzig, B.G. Teubner.
29. Hilbert D. 1918, Axiomatisches Denken, *Mathematische Annalen* **78**: 405–415; in Hilbert 1935, 146–156.
30. Hilbert D. 1922, Neubegründung der Mathematik. Erste Mitteilung, *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität* **1**, 155–177; in Hilbert 1935, 155–177. English translation in Mancosu 1998, 198–214.
31. Hilbert D. 1923, Die logischen Grundlagen der Mathematik, *Mathematische Annalen* **88**: 151–165; in Hilbert 1935, 178–191.
32. Hilbert D. 1926, Über das Unendliche, *Mathematische Annalen* **95**: 161–190. English translation in van Heijenoort 1967, 367–392.
33. Hilbert D. 1928, Die Grundlagen der Mathematik, *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität* **6**: 65–83. English translation in van Heijenoort 1967, 464–479.
34. Hilbert D. 1930, Naturerkennen and Logik, *Naturwissenschaften* **18**: 959–963; in Hilbert 1935, 378–387.
35. Hilbert D. 1931, Die Grundlegung der elementaren Zahlenlehre, *Mathematische Annalen* **104**: 485–494.
36. Hilbert D. 1935, *Gesammelte Abhandlungen* **III**, Berlin, Springer. Reprint New York, Chelsea, 1965.
37. Hilbert D. and Bernays P. 1934, *Grundlagen der Mathematik* **I**, Berlin, Springer-Verlag.
38. Herbrand J. 1968, *Écrits Logiques*, van Heijenoort (ed.), Paris, Presses Universitaires de France.
39. Hintikka J. 2004, On Tarski's Assumptions, *Synthese* **142**: 353–369.
40. Kreisel G. 1958, Mathematical Significance of Consistency Proofs, *The Journal of Symbolic Logic* **23**: 155–182.
41. Kreisel G. 1968, A Survey of Proof Theory, *The Journal of Symbolic Logic* **33**: 321–388.
42. Kreisel G. 1983, Hilbert's Programme, *Dialectica* **12** (1958): 346–372. English expanded version in Benacerraf P. and Putnam H. [1], 207–238.

43. Kreisel G. 1985, Mathematical Logic: Tool and Object lesson for Science, *Synthese* **62**: 139–151.
44. Mancosu P. 1998, *From Brouwer to Hilbert. The Debate on the Foundations of Mathematics in the 1920s*, New York, Oxford University Press.
45. Mancosu P. 1999, Between Russell and Hilbert: Behmann on the Foundations of Mathematics, *The Bulletin of Symbolic Logic* **5**(3): 303–330.
46. Mancosu P. 2005, Harvard 1940–1941: Tarski, Carnap and Quine on a finitistic language of mathematics for science, *History and Philosophy of Logic* **26** (November 2005): 327–357.
47. Mancosu P. 2006, Tarski on Models and Logical Consequence, in Ferreirós J. and Gray J.J. (eds.), *The Architecture of Modern Mathematics*, Chapter 7, Oxford University Press, 2006.
48. Mycielski J. 2004, On the Tension Between Tarski's Nominalism and His Model Theory (Definitions for a Mathematical Model of knowledge), *Annals of Pure and Applied Logic* **126**: 215–224.
49. Poincaré H. 1905, Les mathématiques et la logique, *Revue de Métaphysique et de Morale* **13**: 815–835 and **14**: 17–34.
50. Reid C. 1970, *Hilbert*, New York-Heidelberg-Berlin, Springer-Verlag.
51. Robinson A. 1965, Formalism 64, *Proceedings of the International Congress for Logic, Methodology and Philosophy of Science*, Jerusalemn (1964), Amsterdam, North Holland, 228–246; in H.J. Keisler, S. Körner, W.A.J. Luxemburg and A.D. Young (eds.), *Selected Papers*, New Haven/London, Yale University Press, 1979, 505–523.
52. Robinson A. 1969, From a formalist's point of view, *Dialectica* **23**: 45–49.
53. Rodriguez-Consuegra F. 2005, Tarski's Intuitive Notion of Set, in Sica G. (ed.), *Essays on the Foundations of Mathematics and Logic*, Monza (Italy), Polimetrica International Scientific Publisher, Advanced Studies in Mathematics and Logic, **1**: 227–266.
54. Sieg W. 1999, Hilbet's Programs: 1917–1922, *The Bulletin of Symbolic Logic* **5**(3): 1–44.
55. Skolem T. 1923, Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbaren Veränderlichen mit unendlichem Ausdehnungs-bereich. English translation in van Heijenoort 1967, 302–333.
56. Suppes P. 1988, Philosophical Implications of Tarski's Work, *The Journal of Symbolic Logic* **53**(1): 80–91.
57. Tait W. 1981, Finitism, *The Journal of Philosophy* **78**(9): 487–546.
58. Tarski A. 1930a, Über einige fundamentale Begriffe der Metamathematik, *Compte Rendus de la Société des Sciences et des Lettres de Varsovie* **XXIII**, Cl. **3**: 22–29. Reprinted in Tarski 1986a, **I**, 311–320. English translation in Tarski 1983, 30–37.
59. Tarski A. 1930b, Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I, *Monatshefte für Mathematik und Physik*, **37**: 361–404. Reprinted in Tarski 1986a, **I**, 341–390. English translation in Tarski 1983, 60–109.
60. Tarski A. 1931, Sur les ensembles définissables de nombres réels I, *Fundamenta Mathematicae* **17**: 210–239. Reprinted in Tarski 1986a, **I**, 517–548. English translation in Tarski 1983, 110–142.
61. Tarski A. 1935–36, Grundzüge des Systemenkalküls I, II, *Fundamenta Mathematicae* **25**: 503–526 and **26**: 283–301, in Tarski 1986a, **II**, 25–50 and 223–244. English translation in Tarski 1983, 342–383.
62. Tarski A. 1936a, Der Wahrheitsbegriff in den formalisierten Sprachen, *Studia Philosophica* **I**: 261–405 (first publication in Polish in 1933). English translation in Tarski 1983, 152–278.
63. Tarski A. 1936b, Grundlegung der wisenschaftlichen Semantik, *Actes du Congrès International de Philosophie Scientifique*, Paris, Hermann, 1936, 1–8, in Tarski 1986a, **II**, 259–268. English Translation: The Establishment of Scientific Semantics, *Philosophical and Phenomenological Research* **4** (1944): 341–376, in Tarski 1983, 401–408.
64. Tarski A. 1939/67, *The Completeness of Elementary Algebra and Geometry*, Paris, Institut Blaise Pascal; in Tarski 1986a, **IV**, 289–346.
65. Tarski A. 1948/51, *A Decision Method for Elementary Algebra and Geometry* (prepared for publication by J.C. McKinsey), University of California Press, Berkeley and Los Angeles. In Tarski 1986a, **III**, 297–368.

66. Tarski A. 1944, The Semantic Conception of Truth and the Foundations of Semantics, in Tarski 1986a, **II**, 661–699.
67. Tarski A. 1952, Some Notions and Methods on the Borderline of Algebra and Metamathemtics, *Proceedings of the International Congress of Mathematicians* (Cambridge, Mass. 1950), Providence, American Mathematical Society, 705–720. Tarski 1986a, **III**, 459–476.
68. Tarski A. 1965, *Introduction to logic and to the Methodology of the Deductive Sciences*, New York, Oxford University Press (first edition, 1941).
69. Tarski A. 1983, *Logic, Semantics, Metamathematics*, Papers from 1923 to 1938 translated by J.H. Woodger, Oxforfd, Clarendon Press (first ed. 1956).
70. Tarski A. 1986a, *Collected Papers* **I, II, III, IV**, Givant S.R. and McKenzie R.N. (eds.), Birkhäuser.
71. Tarski A. 1986b, What are Logical Notions, Posthumous Paper edited by Corcoran J., *History and Philosophy of Logic* **7**: 143–154.
72. Tarski A. 1987, A Philosophical Letter of Alfred Tarski, with a prefatory note by Morton White (1944), *Journal of Philosophy* **84**: 1987, 28–32.
73. Tarski A. 1995, Some Current Problems in Metamathematics, Posthumous Paper edited by Tarki J. and Wolenski J., *History and Philosophy of Logic* **16**: 159–168.
74. Tarski A. 2000, Address at the Princeton University Bicentennial Conference on Problems of Mathematics (December 17–19, 1946), Posthumous Paper edited by H. Benis Sinaceur, *The Bulletin of Symbolic Logic* **6**(1): 1–44.
75. Tarski A. and Givant S. 1987, *Formalization of Set Theory without Variables*, American Mathematical Society, Providence, R.I.
76. Van Heijenoort J. 1967, *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, Cambridge (Mass.), Harvard University Press.
77. Webb J.C. 1980, *Mechanism, Mentalism, and Metamathematics*, Dordrecht-Boston-London, D. Reidel Publishing Company.
78. Webb J.C. 1995, Tracking Contradictions in Geometry: The Idea of a Model from Kant to Hilbert, in J. Hintikka (ed.), *From Dedekind to Gödel. Essays on the Development of the Foundations of Mathematics*, Synthese Library, vol. 251, Dordrecht/Boston/London, Kluwer Academic Publishers, 1–20.
79. Weyl H. 1921, Über die neue Grundlagenkrise der Mathematik, *Mathematische Zeitschrift* **10**: 39–79. English translation in Mancosu 1998, 86–118.
80. Weyl H. 1925–27, Die heutige Erkenntnislage in der Mathematik. English translation in Mancosu 1998, 123–142.
81. Weyl H. 1928, Diskussionsbemerkungen zu dem zweitem Hilbertschen Vortrag über die Grundlagen der Mathematik, *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität* **6**: 86–88. English translation in van Heijenoort 1967, 480–484.
82. Wolenski J. 1989, *Logic and Philosophy in the Lvov-Warsaw School*, Synthese Library, vol. 198, Dordrecht/Boston/London, Kluwer Academic Publishers.
83. Wolenski J. 1995, On Tarski's Background, in J. Hintikka (ed.), *From Dedekind to Gödel. Essays on the Development of the Foundations of Mathematics*, Synthese Library, vol. 251, Dordrecht/Boston/London, Kluwer Academic Publishers, 331–342.
84. Wolenski J. 1999, Semantic Revolution. Rudolf Carnap, Kurt Gödel, Alfred Tarski, in Tarski J. and Köhler E. (eds.), *Alfred Tarski and the Vienna Circle. Austro-Polish Connections in Logical Empirism*, Dordrecht/Boston/London, Kluwer Academic Publishers, 1–15.
85. Wolenski J. 2003, Logic, Semantics and Realism, in A. Benmakhlouf (ed.), *Sémantique et épistémologie*, Casablanca, Editions Le Fennec, 2003, 135–148.
86. Zach R. 2003, The Practice of Finitism: Epsilon Calculus and Consistency Proofs in Hilbert's Program, *Synthese* **137**: 211–259.

The Constructive Hilbert Program and the Limits of Martin-Löf Type Theory

Michael Rathjen

1 Introduction

Hilbert's program is [30] one of the truly magnificent projects in the philosophy of mathematics. To carry out this program he founded a new discipline of mathematics, called *"Beweistheorie"*, which was to perform the task of laying to rest all worries about the foundations of mathematics once and for all[1] by securing mathematics via an absolute proof of consistency. The failure of Hilbert's finitist reduction program on account of Gödel's incompleteness results is often gleefully trumpeted. Modern logic, though, has shown that modifications of Hilbert's program are remarkably resilient. These modifications can concern different parts of Hilbert's two step program[2] to validate infinitistic mathematics.

The first kind maintains the goal of a finitistic consistency proof. Here, of course, Gödel's second incompleteness theorem is of utmost relevance in that only a fragment of infinitistic mathematics can be shown to be consistent. Fortunately, results in mathematical logic have led to the conclusion that this fragment encompasses a substantial chunk of scientifically applicable mathematics (cf. [15, 53]). This work bears on the question of the indispensability of set-theoretic foundations for mathematics.

The second kind of modification gives more leeway to the methods allowed in the consistency proof. Such a step is already presaged in the work of the Hilbert school. Notably Bernays has called for a broadened or extended form of finitism (cf. [3]). Rather than a finitistic consistency proof the objective here is to give a constructive and predicative consistency proof for a classical theory T in which large parts of infinitistic mathematics can be developed. In order to undertake such

M. Rathjen (✉)
Department of Pure Mathematics, University of Leeds, Leeds LS2 9JT, UK
e-mail: rathjen@amsta.leeds.ac.uk

This paper is a slightly revised and expanded version of [47].

[1] *"... die Grundlagenfragen einfürallemal aus der Welt zu schaffen."*

[2] The first step being to formalize the whole of mathematics in a formal system T. The second and main step consists in proving the consistency of T.

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_17, © Springer 2005. Previously published in
Synthese 147:1, pp. 81–120.

a study fruitfully one needs to point to a particular formalization of constructive predicative reasoning P, and then investigate whether P is sufficient to prove the consistency of T. The particular framework I shall be concerned with in this paper is an intuitionistic and predicative theory of types which was developed by Martin-Löf. He developed his type theory *"with the philosophical motive of clarifying the syntax and semantics of intuitionistic mathematics"* ([30]). It is intended to be a full scale system for formalizing intuitionistic mathematics. Owing to research in mathematical logic over the last 30 years - the program of reverse mathematics and Feferman's work have been especially instrumental here - one can take a certain fragment of second order arithmetic to be the system T. It turns out that Martin-Löf's type theory P is strong enough to prove the consistency of T, thereby validating infinitistic mathematics and proving a constructive Hilbert program to be feasible. Indeed, the system T alluded to above is so capacious that we do not even know of any result in ordinary mathematics which is not provable in T. Of course, this is a claim regarding ordinary mathematics only; highly set-theoretic topics in mathematics are not amenable to a constructive consistency proof, or, more cautiously put, we do not know how to give constructive consistency proofs for such topics.

The main goal of this paper is to find the limits of Martin-Löf type theory. A demarcation of the latter is important in determining the ultimate boundaries of a constructive Hilbert program. The aim is to single out a fragment of second order arithmetic or classical set theory which encompasses all possible formalizations of Martin-Löf type theory.

Since only a quarter of this paper will actually be concerned with exploring the limits of Martin-Löf type theory, perhaps some words of explanation are in order. The paper was originally written for my PhD students in order to acquaint them with a research area at the interface of proof theory, constructivism, and the philosophy of mathematics that is not readily available in text book form. This accounts for the at times naïve and avuncular diction. The hope, though, is that the paper might be of use to other audiences as well. Among the fields broached here are the areas of proof theory, constructivism, subsystems of arithmetic, reverse mathematics, set theory, Martin-Löf type theory, and the philosophy of mathematics. Several of the aforementioned topics are presented here ab initio with the aim of making them more accessible. The cognoscenti, though, should skim over the first couple of sections and then proceed directly to Sections 5 and 6. The present paper is a slightly expanded version of [47]. At first, I intended to revise the paper substantially, but time constraints did not permit me to do so. I resigned to dilating on Section 6 which is concerned with the limits of Martin-Löf type theory. I also intended to excise some parts (in particular Section 4.3) from [47] which I considered to be embarrassing given that the creators of the theories addressed therein are among the authors of this anthology. But I refrained from that also as I didn't see how to cut out pieces without mutilating the paper or revising it substantially.

The following adumbrates how the paper is organized: In Section 2, fragments of second order arithmetic are introduced and their role for formalizing various parts of ordinary mathematics is discussed. Section 3 surveys different forms of constructivism. Section 4 provides an informal introduction to the ideas underlying

Martin-Löf type theory and also relates them to the Dummett-Prawitz meaning-as-use theory. Section 5 is concerned with subsystems of second order arithmetic which can be shown to be consistent within Martin-Löf type theory. Section 6 is devoted to the limits of Martin-Löf type theory and thus to the limits of a constructive Hilbert program based on it. The final section briefly touches on mathematical statements whose proof depends essentially on the higher infinite in Zermelo-Fraenkel set theory and beyond.

2 Systems for Formalizing Mathematics

A natural modification of Hilbert's program consists in broadening the requirement of reduction to finitary methods by allowing reduction to constructive methods more generally.[3] The objective of our modified constructive Hilbert program is not merely the absence of inconsistency but also the demand for a constructive conception for which there is an absolute guarantee that, whenever one proves a 'real' statement in a sufficiently strong classical theory T, say, a fragment of second order arithmetic or set theory, there would be an interpretation of the proof according to which the theorem is constructively true. Moreover, one would like the theory T to be such as to make the process of formalization of mathematics in T almost trivial, in particular T should be sufficiently strong for all practical purposes. This is a very Hilbertian attitude: show once and for all that non-constructive methods do not lead to false constructive conclusions and then proceed happily on with non-constructive methods.

There are several aspects of a constructive Hilbert program that require clarification. One is to find some basic constructive principles upon which a coherent system of constructive reasoning may be built. Another is to point to a particular framework for formalizing infinitistic mathematics. The latter task will be addressed in this section. It was already observed by Hilbert-Bernays [24] that classical analysis can be formalized within second order arithmetic. Further scrutiny revealed that a small fragment is sufficient. Under the rubric of *Reverse Mathematics* a research program has been initiated by Harvey Friedman some thirty years ago. The idea is to ask whether, given a theorem, one can prove its equivalence to some axiomatic system, with the aim of determining what proof-theoretical resources are necessary for the theorems of mathematics. More precisely, the objective of reverse mathematics is to investigate the role of set existence axioms in ordinary mathematics. The main question can be stated as follows:

---

[3] Such a shift from the original program is implicit in Hilbert-Bernays' [24] apparent acceptance of Gentzen's consistency proof for **PA** under the heading "Überschreitung des bisherigen methodischen Standpunktes der Beweistheorie". The need for a modified Hilbert program has clearly been recognized by Gentzen (cf. [21]) and Bernays [3]: *It thus became apparent that the "finite Standpunkt" is not the only alternative to classical ways of reasoning and is not necessarily implied by the idea of proof theory. An enlarging of the methods of proof theory was therefore suggested: instead of reduction to finitist methods of reasoning it was required only that the arguments be of a constructive character, allowing us to deal with more general forms of inferences.*

Given a specific theorem τ of ordinary mathematics, which set existence axioms are needed in order to prove τ?

Central to the above is the reference to what is called ‘ordinary mathematics’. This concept, of course, doesn’t have a precise definition. Roughly speaking, by ordinary mathematics we mean main-stream, non-set-theoretic mathematics, i.e. the core areas of mathematics which make no essential use of the concepts and methods of set theory and do not essentially depend on the theory of uncountable cardinal numbers. In particular, ordinary mathematics comprises geometry, number theory, calculus, differential equations, real and complex analysis, countable algebra, classical algebra in the style of van der Waerden [56], countable combinatorics, the topology of complete separable metric spaces, and the theory of separable Banach and Frechet spaces. By contrast, the theory of non-separable topological vector spaces, uncountable combinatorics, and general set-theoretic topology are not part of ordinary mathematics. Those parts of mathematics on which set-theoretic assumptions have a strong bearing will be addressed in the last section of this paper.

It is well known that mathematics can be formalized in Zermelo-Fraenkel set theory with the axiom of choice. The framework chosen for studying set existence in reverse mathematics, though, is second order arithmetic rather than set theory. Second order arithmetic, $\mathbf{Z_2}$, is a two-sorted formal system with one sort of variables ranging over natural numbers and the other sort ranging over sets of natural numbers. One advantage of this framework over set theory is that it is more amenable to proof-theoretic investigations. However, at least in my opinion, the particular choice of framework is not pivotal for the program.

For many mathematical theorems τ, there is a weakest natural subsystem $S(\tau)$ of $\mathbf{Z_2}$ such that $S(\tau)$ proves τ. Very often, if a theorem of ordinary mathematics is proved from the weakest possible set existence axioms, the statement of that theorem will turn out to be provably equivalent to those axioms over a still weaker base theory. This theme is referred to as *Reverse Mathematics*. Moreover, it has turned out that $S(\tau)$ often belongs to a small list of specific subsystems of $\mathbf{Z_2}$ dubbed $\mathbf{RCA}_0$, $\mathbf{WKL_0}$, $\mathbf{WKL_0^+}$, $\mathbf{ACA}_0$, $\mathbf{ATR}_0$ and $(\Pi^1_1\mathbf{-CA})_0$, respectively. The systems are enumerated in increasing strength. The main set existence axioms of $\mathbf{RCA}_0$ $\mathbf{ACA}_0$, $\mathbf{ATR}_0$, and $(\Pi^1_1\mathbf{-CA})_0$ are recursive comprehension, arithmetic comprehension, arithmetical transfinite recursion and Π^1_1-comprehension, respectively. Definitions of these systems will be given in the next section. For their role in reverse mathematics see [54].

The theories $\mathbf{WKL_0}$ and $\mathbf{WKL_0^+}$ (defined in Section 2.1) are particularly interesting in pursuing a partial realization of the original Hilbert program. Both theories are of the same proof-theoretic strength as primitive recursive arithmetic, **PRA**, a system which is often considered to be co-extensive with finitism (cf. [55]). The principal set existence axiom of $\mathbf{WKL_0}$ is a non-constructive principle known as *weak König’s lemma* which asserts that any infinite tree of finite sequences of zeros and ones has an infinite path. Friedman [18] proved via model-theoretic methods that $\mathbf{WKL_0}$ is conservative over **PRA** with respect to Π^0_2 sentences.

The question as to which parts of mathematics have applications in science, has also been studied intensively by Feferman (cf. [15, 16]). Over the years he has developed several systems for formalizing mathematics. The system $\mathbf{W}_F$ (cf. [16]) (in honor of H. Weyl) is perhaps the most streamlined. $\mathbf{W}_F$ has flexible finite types (over the natural numbers) and allows for very natural reconstructions of the real and complex numbers (as sets) and much of classical and functional analysis. In $\mathbf{W}_F$ one accepts the completed infinite set of natural numbers as well as classical logic. Though, impredicative set comprehension is taboo. $\mathbf{W}_F$ is conservative over Peano arithmetic.

2.1 Subsystems of Second Order Arithmetic

The purpose of this section is to introduce the formal system of second order arithmetic and several of its subsystems so as to be able to delineate precisely its constructively justifiable parts. Another purpose is to give definitions of the subsystems figuring in reverse mathematics mentioned above.

The most basic system we shall be concerned with is primitive recursive arithmetic, **PRA**, which is a theory about the natural numbers which has function symbols for all primitive recursive functions but in contrast to Peano arithmetic, **PA**, allows for induction only quantifier free formulae.

The language $\mathcal{L}_2$ of second-order arithmetic contains (free and bound) natural number variables $a, b, c, \ldots, x, y, z, \ldots$, (free and bound) set variables $A, B, C, \ldots, X, Y, Z, \ldots$, the constant 0, function symbols $Suc, +, \cdot$, and relation symbols $=, <, \in$. Suc stands for the successor function.

Terms are built up as usual. For $n \in \mathbb{N}$, let $\bar{n}$ be the canonical term denoting n. Formulae are built from the prime formulae $s = t$, $s < t$, and $s \in A$ using $\wedge, \vee, \neg, \forall x, \exists x, \forall X$ and $\exists X$ where s, t are terms.

Note that equality in $\mathcal{L}_2$ is a relation only on numbers. However, equality of sets will be considered a defined notion, namely

$$A = B \quad \text{iff} \quad \forall x[x \in A \leftrightarrow x \in B].$$

The basic axioms in all theories of second-order arithmetic are the defining axioms of 0, $Suc, +, \cdot, <$ and the *induction axiom*

$$\forall X(0 \in X \wedge \forall x(x \in X \to x + 1 \in X) \to \forall x(x \in X)),$$

where $x + 1$ stands for $Suc(x)$. With regard to a collection of $\mathcal{L}_2$ formulae $\mathcal{F}$, the *schema of $\mathcal{F}$-induction* consists of the formulae

$$\mathbf{IND}_{\mathcal{F}} \qquad \varphi(0) \wedge \forall x(\varphi(x) \to \varphi(x+1)) \to \forall x \varphi(x),$$

where φ belongs to $\mathcal{F}$. If $\mathcal{F}$ is the collection of all $\mathcal{L}_2$ formulae we denote the schema by **IND** rather than $\mathbf{IND}_{\mathcal{F}}$. Note that $\mathbf{IND}_{\Delta_0^0}$ implies the induction axiom.

The strength of systems of second order arithmetic is largely owed to *comprehension schemes,* which assert, roughly speaking, that if we specify a collection X of numbers by a formula φ of a particular type then X is a set. We consider the axiom schema of $\mathcal{C}$*-comprehension* for formula classes $\mathcal{C}$ which is given by

$$\mathcal{C} - \mathbf{CA} \qquad \exists X \forall u (u \in X \leftrightarrow \varphi(u))$$

for all formulae $\varphi \in \mathcal{C}$ in which X does not occur.

The fundamental idea of reverse mathematics is to gauge the proof-theoretical strength of a mathematical theorem by classifying how much comprehension is needed to establish the existence of the sets needed to prove the theorem. That is, we "reverse" the theorem to derive some sort of comprehension scheme. The gauging measure is that of the allowable "logical complexity" of the φ's. Typically, this complexity might be the allowable quantifier depth of φ.

Numerical quantifiers are called bounded if they occur in the context $\forall x(x < s \rightarrow \ldots)$ or $\exists x(x < s \wedge \ldots)$ for a term s which does not contain x. The Δ^0_0-formulae are those formulae in which all quantifiers are bounded numerical quantifiers. For instance, the formula asserting that "x is a prime number" is a Δ^0_0-formula, i.e., $Prime(x) \equiv \forall u < x + 1 \, \forall v < x + 1 (u \cdot v = x \rightarrow u = 1 \vee v = 1)$, where 1 is $Suc(0)$.

For $k > 0$, Σ^0_k-formulae are formulae of the form $\exists x_1 \forall x_2 \ldots Q x_k \varphi$, while Π^0_k-formulae are those of the form $\forall x_1 \exists x_2 \ldots Q x_k \varphi$, where φ is Δ^0_0 and the numerical quantifiers alternate in each of the prefixes. The union of all Π^0_k- and Σ^0_k-formulae for all $k \in \mathbb{N}$ is the class of *arithmetical* or Π^0_∞*-formulae*. The superscript "0" refers to the fact that there are no set quantifiers. We obtain a similar hierarchy if we allow set quantification by putting a superscript "1" and counting the number of alterations of set quantifiers over an arithmetical matrix. The Σ^1_k-formulae (Π^1_k-formulae) are the formulae $\exists X_1 \forall X_2 \ldots Q X_k \varphi$ (resp. $\forall X_1 \exists X_2 \ldots Q x_k \varphi$) for arithmetical φ, where the set quantifiers alternate in each of the prefixes.

For each axiom schema **Ax** we denote by (**Ax**) the theory consisting of the basic arithmetical axioms, the schema of induction **IND**, and the schema **Ax**. If we replace the schema of induction by the induction axiom, we denote the resulting theory by $(\mathbf{Ax})\restriction$.

An example for these notations is the theory ($\Pi^1_1 - \mathbf{CA}$) which contains the induction schema, whereas $(\Pi^1_1 - \mathbf{CA}) \restriction$ contains only the induction axiom in addition to the comprehension schema for $\mathbf{P}^1_1$-formulae.

In the framework of these theories one can introduce defined symbols for all primitive recursive functions. In particular, let $\langle , \rangle : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ be a primitive recursive and bijective pairing function.

The *xth* section of U is defined by $U_x := \{y : \langle x, y \rangle \in U\}$. Observe that a set U is uniquely determined by its sections on account of $\langle , \rangle$'s bijectivity.

Any set R gives rise to a binary relation $\prec_R$ defined by $y \prec_R x := \langle y, x \rangle \in R$.

Using the latter coding, we can formulate the axiom of choice for formulae φ in $\mathcal{C}$ by

$$\mathcal{C} - \mathbf{AC} \qquad \forall x \exists Y \varphi(x, Y) \to \exists Y \forall x \varphi(x, Y_x).$$

A special form of comprehension is Δ^1_n-comprehension, that is

$$\Delta^1_n - \mathbf{CA} \qquad \forall u\big[\varphi(u) \leftrightarrow \vartheta(u)\big] \to \exists X \forall u(u \in X \leftrightarrow \varphi(u))$$

for all Π^1_n-formula φ and Σ^1_n-formula ϑ. Δ^0_n-comprehension is defined by requiring that φ and ϑ are Π^0_n-formulae and Σ^0_n-formulae, respectively.

In set theory one has the principle of set induction which says that whenever a property propagates from the elements of any set to the set itself, then all sets have the property. In the context of second order arithmetic the equivalent of set induction is the schema of *transfinite induction*

$$\mathbf{TI} \qquad \forall X\big[\mathrm{WF}(\prec_X) \ \wedge \ \forall u\big(\forall v \prec_X u \varphi(v) \to \varphi(u)\big) \ \to \ \forall u \varphi(u)\big]$$

for all formulae φ, where $\mathrm{WF}(\prec_X)$ expresses that $\prec_X$ is well-founded, i.e., that there are no infinite descending sequences with respect to $\prec_X$. Classically $\mathrm{WF}(\prec_X)$ is equivalent to

$$\forall Y \ [\forall u(\forall v \prec_X u \ v \in Y \ \to \ u \in Y) \ \to \ \forall u \ u \in Y.]$$

We have now introduced the schemes for defining the preferred systems of reverse mathematics. $\mathbf{RCA}_0$ is the theory $(\Delta^0_1 - \mathbf{CA}){\restriction} \ +\mathbf{IND}_{\Sigma^0_1}$. $\mathbf{ACA}_0$ denotes the system $(\Pi^0_\infty - \mathbf{CA}){\restriction}$ and $\mathbf{ATR}_0$ is $\mathbf{ACA}_0$ augmented by a schema which asserts that Π^0_1-comprehension (or the Turing jump) may be iterated along any well-ordering. $(\Pi^1_1{-}\mathbf{CA})_0$ denotes $(\Pi^1_1{-}\mathbf{CA}){\restriction}$. The principal set existence axiom of $\mathbf{WKL_0}$ is an extension of $\mathbf{RCA}_0$ by *weak König's lemma* which asserts that any infinite tree of finite sequences of zeros and ones has an infinite path. The mathematically stronger system $\mathbf{WKL_0^+}$ was defined by Brown and Simpson (cf. [7, 8]). Let $2^{<\mathbb{N}}$ denote the set of finite sequences of zeros and ones. The axioms of $\mathbf{WKL_0^+}$ are those of $\mathbf{WKL_0}$ plus a scheme which amounts to a strong formal version of the Baire category theorem for the Cantor space $2^{\mathbb{N}}$. More formally, the additional scheme expresses that, given a sequence of dense subcollections of $2^{<\mathbb{N}}$ which is arithmetically definable from a given set, there exists an infinite sequence of zeros and ones which meets each of the given dense subcollections. The advantage of $\mathbf{WKL_0^+}$ over $\mathbf{WKL_0}$ is that $\mathbf{WKL_0^+}$ proves several important theorems of functional analysis which are apparently not provable in $\mathbf{WKL_0}$. Brown and Simpson used forcing to prove that $\mathbf{WKL_0^+}$ is still Π^0_2 conservative over $\mathbf{PRA}$.

2.2 How much of Second Order Arithmetic is Needed?

More precisely the foregoing question asks which part of second order arithmetic is needed for carrying out ordinary mathematics. The program of reverse mathematics as well as Feferman's investigations have amassed a large body of detailed results which allows one to draw the conclusion that a little bit goes a long way. Simpson [53] estimates “that at least 85% of existing mathematics can be formalized in $\mathbf{WKL_0}$ or $\mathbf{WKL_0^+}$ or stronger systems which are conservative over **PRA** with respect to Π^0_2 sentences.” Similarly, Feferman conjectures that the overwhelming part of scientifically applicable mathematics can be formalized in systems (like $\mathbf{W}_F$) which are conservative over Peano arithmetic.

The focus of this section, however, is rather on the other end of the spectrum, where one is interested in mathematical theorems which encapsulate consistency strength beyond **PRA** and Peano arithmetic. Since 1931, the year Gödel’s Incompleteness Theorems were published, logicians have been looking for a strictly mathematical example of an incompleteness in first-order Peano arithmetic, one which is mathematically simple and interesting and does not require the numerical coding of notions from logic. The first such examples were found early in 1977. The most elegant of these is a strengthening of the Finite Ramsey Theorem due to Paris and Harrington (cf. [37]). The original proofs of the independence of combinatorial statements from **PA** all used techniques from non-standard models of arithmetic. Only later on alternative proofs using proof-theoretic techniques were found, though results from ordinal-theoretic proof theory turned out to be pivotal in providing independence results for theories stronger than **PA**, and even led to a new combinatorial statement. The stronger theories referred to are Friedman’s system $\mathbf{ATR}_0$ of *arithmetical transfinite recursion* and the system $(\Pi^1_1 - \mathbf{CA})\restriction$ based on Π^1_1-comprehension. The independent combinatorial statements have their origin in certain embeddability questions in the theory of finite graphs. The first is a famous theorem of Kruskal asserting that every infinite set of finite trees has only finitely many minimal elements. It turned out that Kruskal’s theorem is not provable in $\mathbf{ATR}_0$ (see [52]). An even more striking example of this independence phenomenon is provided by the *graph minor theorem*, GMT, of Robertson and Seymour which is one of the most important theorems of graph theory (see [10]). It was shown by Friedman, Robertson, and Seymour [19] that the GMT is not provable in $(\Pi^1_1 - \mathbf{CA})\restriction$. Actually, the original proof of the GMT for graphs of bounded tree width used Friedman’s extended version of Kruskal’s theorem with the “gap condition”, which was gleaned from the proof-theoretic ordinal analysis of $(\Pi^1_1 - \mathbf{CA})\restriction$ and specifically designed to construct stronger incompletenesses in Peano arithmetic as part of the reverse mathematical program, so that the metamathematical considerations had a considerable mathematical spin-off.

Looking for an upper bound, it has been claimed that the system $(\Pi^1_1 - \mathbf{CA}) + \mathbf{TI}$ suffices for proving the graph minor theorem. Indeed, up till now no statements of everyday mathematics have been found that require more than means available in $(\Pi^1_1 - \mathbf{CA}) + \mathbf{TI}$.[4]

[4] However, there are certain mathematical statements whose proof requires the consistency strength of large cardinals. This topic will be briefly touched upon in the last section.

3 Forms of Constructivism

In the foregoing section we presented a framework sufficient for the needs of ordinary infinitistic mathematics. Finding a suitable system for formalizing constructive mathematics presents a more difficult task. Historically there have been differing "schools" of constructivism some of which are mutually incompatible. However, at least in my opinion, Martin-Löf's theory is the most thoroughly worked out system and, I think, philosophically the most convincing candidate. After recalling some history of constructivism in this section, the next section will provide an informal introduction to the ideas underlying Martin-Löf type theory.

One of the first facts that has to be taken into account is that 'constructive mathematics' is not a single, clearly defined body of mathematics. Indeed, several different brands of constructivism can be discerned:

1. Predicativism (Poincaré, Russell, Weyl, Lorenzen)
2. Intuitionism (Brouwer, Heyting, et. al.)
3. Russian constructivism (Markov, Shanin)
4. Bishop's constructive mathematics

Predicativism took shape in the writings of Poincaré and Russell in response to the paradoxes. Russell discerned the common underlying root for the paradoxes 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.* ([48], p. 224)

Thus Russell chimes in with Poincaré's anathemizing of so-called *impredicative definitions*. 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* ([57], p. 243).

In his book *Das Kontinuum*, Weyl initiated a predicative approach to the the real numbers and gave a viable account of a substantial chunk of analysis. What are the ideas and principles that his "predicative view" is grounded in? 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. As 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 in an ascending ladder of systems $RA_0, RA_1, \ldots, RA_\alpha, \ldots$ was first presented in Kreisel [28] and than taken up by Schütte and Feferman 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.

Brouwer regarded *intuitionistic* mathematics as the only legitimate form of mathematics, viz. the only kind of mathematics which subjectively could be described as "indubitable". Mathematics, in his view, consists of mental constructions, performed by the individual in free action. In the main, mathematics is a languageless and solitary activity, where words merely accompany mathematical "constructions originating by the self-unfolding of the primordial intuition ..." ([5]). In Brouwerian intuitionism, much attention is given to the structure of the continuum. Infinite proceeding sequences and other concepts, such as choice sequences are introduced. Consideration of the nature of these sequences as perceived by the ideal mathematician leads one to the conclusion that every operation from $\mathbb{N}^{\mathbb{N}}$ to $\mathbb{N}$ is continuous. The latter is known as *Brouwer's principle* (for natural numbers), BP_0. The pivotal consequence of BP_0 is that all functions from the reals to the reals are continuous. As a result, one arrives in Brouwer's form of constructivism at theorems which contradict classical mathematics.

The concept of algorithm or recursive function is fundamental to the *Russian schools* of Markov and Shanin. Contrary to Brouwer, this school takes the viewpoint that mathematical objects must be concrete, or at least have a constructive description, as a word in an alphabet, or equivalently, as an integer, for only on such objects do recursive functions operate. Furthermore, Markov adopts what he calls *Church's thesis*, CT, which asserts that whenever we see a quantifier combination $\forall n \in \mathbb{N}\, \exists m \in \mathbb{N}\, A(n, m)$, we can find a recursive function f which produces m from n, i.e. $\forall n \in \mathbb{N}\, A(n, f(n))$. On the other hand, as far as pure logic is concerned he augments Brouwer's intuitionistic logic by what is known as *Markov's principle*, MP, which may be expressed as

$$\forall n \in \mathbb{N}\, [A(n) \vee \neg A(n)] \ \wedge\ \neg \forall n \in \mathbb{N}\, \neg A(n) \ \rightarrow\ \exists n \in \mathbb{N}\, A(n),$$

with A containing natural number parameters only. The rationale for accepting MP may be phrased as follows. Suppose A is a predicate of natural numbers which can be decided for each number; and we also know by indirect arguments that there should be an n such that $A(n)$. Then a computer with unbounded memory could be programmed to search through $\mathbb{N}$ for a number n such that $A(n)$ and we should be convinced that it will eventually find one. As an example for an application of MP to the reals one obtains $\forall x \in \mathbb{R}\,(\neg x \leqslant 0 \rightarrow x > 0)$.

In 1967 Bishop published his *Foundations of constructive analysis* [4] in which he carried out an informal development of constructive analysis which went substantially further mathematically than anything done before by constructivists. Bishop was a confirmed constructivist, as was Brouwer. However, what was novel about Bishop's work was that it could be read as a piece of classical mathematics as well. Bishop works with informal notions of constructive function and set. Among the constructivists schools, his standpoint uses the fewest assumptions in the mathematics. In fact, letting BC stand for Bishop-style constructivism, one may characterize the other philosophies roughly as follows:

- Russian constructivism = $BC + MP + CT$
- Brouwerian intuitionism = $BC + BP_0 + BI$,

where BI is bar induction, which is a form of transfinite induction on well-founded trees. With classical logic, BI is equivalent to the principle **TI** defined in Section 2.1.[5]

Several frameworks for constructivism that relate to Bishop's constructive mathematics as theories like **ZFC** relate to Cantorian set theory have been proposed by Myhill, Martin-Löf, and Feferman. Myhill (cf. [33]) developed his constructive set theory with the aim of isolating the principles underlying Bishop's conception of what sets and functions are. Moreover, he wanted "these principles to be such as to make the process of formalization completely trivial, as it is in the classical case" ([33], p. 347). Indeed, while he uses other primitives in his set theory CST besides the notion of set, it can be viewed as a subsystem of **ZF**. The advantage of this is that the ideas, conventions and practise of the set theoretical presentation of ordinary mathematics can be used in the set theoretical development of constructive mathematics, too. Feferman's *Explicit Mathematics* is a theory of operations and classes ([13, 14]), T_0, which is suitable for representing Bishop-style constructive mathematics as well as generalized recursion, including direct expression of structural concepts which admit self-application.

The intuitionistic theory of types **MLTT** as developed by Martin-Löf (cf. [29, 31]) is also intended to be a system for formalizing intuitionistic mathematics. However, Martin-Löf probes far deeper in several respects. Not only has he developed a formal system **MLTT** but also a philosophical underpinning of constructivism which is encapsulated in his informal semantics for **MLTT**, called "meaning explanation". The latter is a systematic method of assigning meaning to the assertions

---

[5] This is the reason why the principle **TI** is often referred to as *bar induction*.

of **MLTT** which enables him to justify the rules of **MLTT** by showing their validity with respect to that semantics. It is perhaps not an exaggeration to say that Martin-Löf's theory gives rise to a full scale philosophy of constructivism.

Type theory is a logic free theory of constructions within which the logical notions can be defined whereas systems of explicit mathematics and constructive set theory leave the logical notions unanalyzed. For this reason I consider type theory philosophically more fundamental.[6]

Integral to an understanding of Martin-Löf's theory of types, notably his justification of the logical laws, is an awareness of the distinction between the notions of *judgement*[7] and *proposition*. The point of view he adopts is that logical operations (constants) operate on propositions whereas the logical laws (rules of inference) operate on judgements. Performing the logical operation $\vee$ leads from propositions A and B to a new proposition $A \vee B$. By contrast, the premises and conclusion of a logical inference are always judgements. In the order of conceptual priority the notion of judgement comes before the notion of proposition since the assertion *A is a proposition* is a judgement.

The notion of proposition is a semantic notion. In a first approach, a proposition could be construed as a meaningful statement describing a state of affairs. Traditionally, a proposition is something that is either true or false. In the case of mathematical statements involving quantifiers ranging over infinite domains, however, by adopting such a view one is compelled to postulate a transcendent realm of mathematical objects which determines their meaning and truth value. Like Brouwer, Martin-Löf repudiates such an account as a myth. He explains that the meaning of a mathematical statement is not independent of our cognitive activity, that it is subject related, relative to the knowing subject in Kantian terminology. Kolmogorov observed that the laws of the intuitionisic propositional calculus become evident upon conceiving propositional variables as ranging over problems or tasks.[8] In a similar vein, Martin-Löf explains the notion of proposition as follows:

> *Returning now to the form of judgement* 'A is a proposition'*, the semantical explanation which goes together with it is this, and here I am using the knowledge theoretical formulation, that to know a proposition, which may be replaced, if you want, by problem, expectation or intention, you must know what counts as a verification, solution, fulfillment, or realization of it. ([32], p. 34)*

---

[6] Incidentally, this view was shared by Myhill: *Since completing this paper I have become familiar with some recent unpublished work of Martin-Löf, ... While the complexity of Martin-Löf's system makes extremely unlikely its general adoption as a definitive formalization of constructive mathematics, it pushes its analysis far deeper than we do, ... In any case no further work should be done in this area without familiarity with Martin-Löf's work, whose speedy publication we anticipate with pleasure. ([33], p. 355)*

[7] The word "judgement" is closely related to the German word "Urteil". Urteil was the central notion of logic in Kant's philosophy. In his transcendental logic, which forms part of the *Kritik der reinen Vernunft*, Kant arrives at his categories by discerning the various forms of judgements.

[8] He used the German word *Aufgabe*.

In keeping with the above meaning explanation of the judgement *'A is a proposition'*, to make the judgement *'A is true'* you must have knowledge how to verify A, and for a mathematical proposition A a method to verify A is nothing else but a proof of A.

The most forceful criticism of the idea of a knowledge transcendent notion of mathematical truth has been put forward by Dummett. His arguments will be related in Section 4.3.

4 Martin-Löf's Theory of Types

When we talk about Martin-Löf type theory we refer to more than just a formal system as the meaning explanations for the rules form an essential ingredient of it. The language with which we will be concerned here is a full scale system which accounts for essentially everything one does in mathematics. The origins of this system can be traced to the natural deduction systems of Gentzen, taken in conjunction with Prawitz's reduction procedures, and to Gödel's Dialectica system. Moreover, **MLTT** is an open ended framework in that one may always add new types and rules providing they are amenable to a meaning explanation which validates them. In point of fact, a particular powerful way of going beyond an existing formalization T of **MLTT** is by reflection about T. This is formally mirrored by the introduction of universes into **MLTT**.

Distinctive features of **MLTT** are the following:

- *meaning explanations* for rules
- the full use of the *propositions-as-types* paradigm to represent logic
- strength and expressiveness are obtained through the use of *inductive data types* and *reflection*, i.e. type universes.

These features will be addressed one by one in what follows.

4.1 A First Glimpse of Type Theory

Martin-Löf's formal language has a system of rules for deriving judgements. This is in contrast to the standard formal systems which involve rules for deriving propositions. The distinction between propositions and judgements is essential for Martin-Löf. What we combine by means of the logical operations ($\rightarrow, \wedge, \vee, \neg, \forall, \exists$) and hold true are propositions. When we hold a proposition to be true, we make a judgement. The fundamental notions of type theory are introduced in the four forms of judgement:

1. A is a type (abbr. A *type*),
2. A and B are equal types (abbr. $A = B$),
3. a is an element of type A (abbr. $a : A$),
4. a, b are equal elements of type A (abbr. $a = b : A$).

There is a qualitative distinction between arbitrary elements of a type and canonical elements of a type. A type A is defined by stating what we have to do in order to construct a *canonical* element of the type and what conditions need to be satisfied for such canonical elements to be equal. By contrast, an arbitrary element of type A is a method or program which when executed yields a canonical element of type A. Two arbitrary elements a, b of A are equal if, when executed, a and b evaluate to equal canonical elements of A. Equality between canonical elements of the same type should be reflexive, symmetric, and transitive. For each type there are formation rules for that type and introduction rules to construct canonical elements of that type. These are best illustrated by means of a concrete example. For ease of presentation we will neglect all rules dealing with equality. The type of natural number is introduced by the following rules:

$$(\mathbb{N}\text{-formation}) \qquad \mathbb{N}\ type$$

$$(\mathbb{N}\text{-introduction}) \qquad 0 : \mathbb{N} \qquad \frac{a : \mathbb{N}}{Sa : \mathbb{N}}$$

The introduction rules state how canonical elements of $\mathbb{N}$ are generated. On the other hand, $2 + 2$ should be regarded as an element of type $\mathbb{N}$ as well, since we can evaluate it to a canonical element of type $\mathbb{N}$. But obviously $2 + 2$ is not obtainable from the given rules, viz. the judgement $2 + 2 : \mathbb{N}$ is not derivable via $\mathbb{N}$-introduction. Martin-Löf regards the expression, "$2 + 2$" as a program, which gives instructions for its own evaluation. In its evaluated form (in his theories) it will be the canonical element $SSSS0$. Therefore we shall need further rules, called *elimination rules*, which allow one to derive judgements of the form $t : A$ for non-canonical t. However, before stating any elimination rules let us study some further forms of types and their canonical elements, called canonical elements henceforth.

$$A \times B \qquad A \to B \qquad (\Pi x : A)B(x) \qquad A + B \qquad (\Sigma x : A)B(x).$$

A canonical element of the Cartesian product type $A \times B$ is a pair, and can be written as (a, b), where $a : A$ and $b : A$. The pertaining rules are:

$$(\times\text{-introduction}) \qquad \frac{a : A \qquad b : B}{(a, b) : A \times B} \qquad \frac{a = a' : A \qquad b = b' : B}{(a, b) = (a', b') : A \times B}$$

A canonical element of $A \to B$ is an expression $(\lambda x)t(x)$ which (informally) denotes a function given by a term $t(x)$ such that $t(a)$, the result of substituting a for x, is an element of B for each $a : A$. In the natural deduction calculus, ($\to$-introduction) takes the forms

$$\frac{\begin{array}{c}[x : A]\\ t(x) : B\end{array}}{(\lambda x)t(x) : A \to B} \quad \frac{\begin{array}{c}[x : A]\\ t(x) = s(x) : B\end{array}}{(\lambda x)t(x) = (\lambda x)s(x) : A \to B},$$

where $[x : A]$ indicates an assumption which gets discharged by the inference.

Suppose A is a type and $B(a)$ is a type for each $a : A$. Then we can form a new type $(\Pi x : A)B$ whose canonical elements are of the form $(\lambda x)t(x)$ where $t(a)$ is an element of B for each $a : A$.

The canonical elements of $A + B$ are of the forms $i(a)$ and $j(b)$, where $a : A$ and $b : B$, respectively.

If $B(a)$ is a type for each $a : A$, then $(\Sigma x : A)B(x)$ is a type whose canonical elements are pairs (a, b) with $a : A$ and $b : B(a)$.

There is a certain pattern for forming canonical elements in all the above cases. Each element in canonical form is built from its components using special *constructors*. $\mathbb{N}$ has the constructors 0 and S; $A \to B$ has the constructor λ; $A + B$ has the constructors i and j; and so forth.

The dual of the introduction rules for obtaining canonical elements of a type are its *elimination rules*. The elimination rules for a type A are, as it were, natural consequences of its introduction rules. They amount to saying that all canonical elements of A are generated by exactly the means laid down in the introduction rules, viz. there are no other ways to form canonical elements. Martin-Löf is expanding here on Gentzen's ideas of dividing logical rules into introduction and elimination rules. Referring to the logical constants, Gentzen [20] explains the harmony between these two kinds of rules as follows: *"an introduction rule gives, as it were, a definition of the constant in question"* while *"an elimination rule is only a consequence of the corresponding introduction rule, which may be expressed somewhat as follows: when making an inference by an elimination rule, we are allowed to 'use' only what the principal sign of the major premiss 'means' according to the introduction rule for this sign."*

In the case of the type of natural numbers $\mathbb{N}$ the elimination rules amount to familiar rules for structural induction and recursion over $\mathbb{N}$. The formal rendering of elimination rules for an arbitrary type requires a new constant. Associated with each type will be a *selector* (dual to the notion of a constructor, above), given as an implicitly defined constant, whose defining equations express the elimination rules for the type. As the elimination rules for $\mathbb{N}$ are rather involved (but see Section 4.4 below) let us study the much simpler case of the type $A \to B$. Here the rules are:[9]

[9] Martin-Löf actually breaks up the rules which we collectively call elimination rules, into two groups dubbed elimination and equality rules, respectively. His elimination rules explain how non-canonical elements are formed via the selector, and the equation rules explain how the selector operates on the canonical elements.

(→ -elimination) $$\frac{c : A \to B \quad a : A}{\mathbf{App}(c, a) : B}$$

$$\frac{c = d : A \to B \quad a = b : A}{\mathbf{App}(c, a) = \mathbf{App}(d, b) : B}$$

$$\frac{\begin{array}{c}[x : A] \\ t(x) : B\end{array} \quad a : A}{\mathbf{App}((\lambda x)t(x), a) = t(a) : B}$$

$$\frac{c : A \to B}{c = (\lambda x)\mathbf{App}(c, x) : A \to B}$$

To explain the meaning of the selector **App** suppose $c : A \to B$ and $a : A$. Then $\mathbf{App}(c, a)$ is a method for obtaining a canonical element of B which is executed as follows. c is a method which yields a canonical element $(\lambda x)t(x)$ of $A \to B$. Substituting a for x leads to $t(a) : B$. Thus, through evaluating $t(a)$ we finally arrive at a canonical element of B.

4.2 The Proposition-as-types Interpretation

So far we have avoided the question how standard logical operations are to be treated in type theory. It turns out that the first glimpse of type theory given above suffices for that task. The point of view to be adhered to is that propositions are types and that logical operations on propositions correspond to the appropriate type forming operations in line with the *Brouwer-Heyting-Kolmogorov interpretation* (BHK-interpretation) of the constructive meaning of the logical constants, whereby each proposition is identified with the type of its proofs.[10] For example the constructive meaning of an implication $A \supset B$ consists in saying that a proof of $A \supset B$ is a constructive procedure that transforms each proof of A into a proof of B. It seems that the notion of constructive procedure used here must be taken as a primitive notion. The table below gives a dictionary for translating logical operations into type theory.

Logical notion	**Type-theoretic notion**
proposition	type
proof of A	element of A
A is true	A has an element
$A \wedge B$	$A \times B$

[10] On the formal level, the analogy between propositions and types was discovered by Curry and Feys [9] and further developed by Howard [26].

$A \supset B$	$A \to B$
$A \vee B$	$A + B$
$(\forall x{\in}A)B(x)$	$(\Pi x : A)B(x)$
$(\exists x{\in}A)B(x)$	$(\Sigma x : A)B(x)$

Note that the treatment of quantifiers in **MLTT** strictly adheres to Brouwer's dictum that quantified variables should range only over already-defined sets/types A.

In rendering propositions as types, the elements of a proposition are to be understood as proofs and thus its canonical elements could be termed *canonical proofs*. The inductive nature of the canonical proofs can be made more explicit by the following table:

a canonical proof of	**has the form**
$A \wedge B$	(a, b), where a is a proof of A and b is a proof of B
$A \supset B$	$(\lambda x)t(x)$, where $t(a)$ is a proof of B whenever a is a proof of A
$A \vee B$	$i(a)$, where a is a proof of A, or $j(b)$, where b is a proof of B
$(\forall x{\in}A)B(x)$	$(\lambda x)t(x)$, where $t(a)$ is a proof of $B(a)$ whenever a is an element of A
$(\exists x{\in}A)B(x)$	(a, b), where a is an element of A and b is a proof of $B(a)$

4.3 The Dummett-Prawitz Meaning-as-use Theory and MLTT

Inspired by ideas of Wittgenstein, Dummett (cf. [11, 12]) has brought forward philosophical arguments against a platonistic theory of meaning, amounting to a rejection of classical logic in favour of intuitionistic logic. In order for such a rejection to be conclusive, Prawitz [38] has expanded on Dummett's ideas who suggested that a theory of meaning formulated in terms of proofs or rules also ought to take note of Gentzen's fundamental insight that a correspondence or duality obtains between the rules for asserting a sentence and the rules for drawing consequences from it. In many aspects the Dummett-Prawitz theory of meaning-as-use is closely related to Martin-Löf's understanding of the meaning of the logical operators and thus I consider it useful to intersperse a brief account of their theory here. On the one hand this might be instructive and illuminating for an understanding of certain aspects of Martin-Löf's meaning explanations. On the other hand, it seems that **MLTT** overcomes a deficiency of the Dummett-Prawitz semantics which is due to the latter's rather narrow focus on logical reasoning. In mathematics, logic retreats into the background and mathematical objects and constructions occupy center stage. The interrelationship between logical inferences and mathematical constructions connects together logic and mathematics. Logic gets intertwined with mathematical objects and operations, and it appears that its role therein cannot be separated from mathematical constructions. Indeed, not only does Martin-Löf's work demonstrate that a coherent theory of meaning can be developed along the lines of Dummett-Prawitz for a rich and elaborate part of mathematics but also that logical operators can be construed as special cases of more general mathematical operations.

Below I shall not undertake any detailed analysis of all the facets of of Dummett's and Prawitz's arguments and confine myself to a rough sketch of the main points.

4.3.1 The Meaning-as-Use Thesis

The classical/platonist mathematician holds that the meaning of a mathematical statement is determined by its truth conditions in the (abstract) realm of mathematical objects existing independently of us. This view, however, appears to be very detached from mathematical practice. If we look back at the way that mathematics is actually taught, it seems that what we learn is not to establish truth-conditions of sentences in a transcendent world but rather what is to be counted as establishing the truth of sentences, that is to say, we are trained in the art of proving mathematical sentences. And it therefore seems that the invocation of this transcendence serves merely as an *Überbau* which fails to add anything to the elucidation of what it is to know a mathematical statement.

Dummett argues that meaning cannot be separated from use and must be recognizable by us. His tenet is that the meaning of a sentence must be fully manifest in its use, where use is to be taken in a very broad sense covering all its aspects. Use exhaustively determines meaning in the sense that two expressions which are always used in the same way ought to have the same meaning. The main lines of support for Dummett's thesis are the following:

> *... that meaning has to be communicable and that communication has to be observable: to assume that there is some ingredient in the meaning of a sentence which cannot become manifest in the use made of it is to assume that part of the meaning cannot be communicated. This part of the meaning would then be irrelevant when the statement was used in communication. Such a meaning would therefore be of no importance for mathematics understood as a social enterprise in which many people can cooperate and exchange results with each other.* ([38], p. 4)

1. When we learn a language in a primordial way, i.e. not by translation into another language, all we learn is how to use expressions correctly. The grasp of the meaning of a sentence in such a language must therefore consists in our ability to use it correctly.
2. Knowledge of the meaning of an expression can sometimes be demonstrated in the manner of Socrates by explicitly defining it in terms of other expressions of which the meaning is already known. However, if there is not to be an infinite regress, the meaning of a statement cannot consist solely of explicit verbalizable knowledge and thus knowledge of meaning must ultimately be traced back to

> implicit knowledge. Implicit knowledge of the meaning of a sentence A can only manifest itself in the ability to use it or to respond to its use by others in a certain observable way. Thus the only way of acquiring such knowledge is by observing and learning its use. Therefore the meaning of a sentence cannot transcend its total use. In short: We know the meaning of A if we know under what conditions A may be correctly asserted.

Specialized to the case of mathematics, the above views on the meaning of a sentence amount to agreeing with the intuitionists that meaning in mathematics has somehow to be understood in terms of proofs.

4.3.2 Molecular Versus Holistic Views on Meaning

The principle that use determines meaning does not rule out the possibility that the meaning of a single statement can only be understood with regard to the framework of language as a whole. An extreme holistic view claims "that nothing less than the total use of the language determines the meaning of an individual sentence." ([38], p. 7) On such a view, it is not possible to develop a meaning theory which explains the meaning of single sentences in terms of their constituent parts or even to adhere to a milder form of holism where one singles out a privileged class of sentences which are amenable to such a molecular meaning explanation and which then gives meaning to other sentences by their deductive relationships with this class. Dummett rejects the drastic form of holism on the grounds that meaning must be recognizable by us through use in particular situations and 'total use of language' blatantly escapes recognizability.

A prominent example of a restricted form of holism in mathematics is, of course, underlying Hilbert's program, where the privileged sentences are called 'real sentences'. Partial holism is also a common standpoint in the philosophy of science. Here one distinguishes between theoretical and observational sentences. The latter are, by and large, endowed with a meaning in isolation whereas the meaning of the former is understood as determined by their role in deducing and refuting privileged observational sentences.

Dummett, however, repudiates all kinds of holistic meaning theories and maintains a molecular view. Following Frege, he sees the meaning of a sentence as being determined by the way it is built up from its constituent parts. Thus a sentence carries an individual meaning which is generated from its atomic components (having immediate meaning) via logical operations.

4.3.3 Meaning-as-Use and the BHK-Interpretation

The meaning-as-use thesis asserts that the meaning of a statement is determined by the conditions under which it may be correctly asserted. Turning to a mathematical statement A, its meaning could be explained by what it means to give a proof of A or what counts as a proof of A. The BHK-interpretation mentioned above provides a precise rendering of what counts as a proof of a mathematical proposition. It also adheres to the molecular semantical view in that the proofs of a complex proposition are explained in terms of proof conditions for its components. To distinguish the notion of proof supported by the BHK-interpretation from others, let us call them *canonical proofs* or *direct proofs*. A first attempt at a meaning theory for mathematical statements A could be framed as follows:

(MT_1) To know the meaning of A is to know the conditions for asserting A.
The condition for asserting A is to know a canonical proof of A.

Under the proposition-as-types view, the canonical proofs of a proposition are precisely the ones generated by the introduction rules for the pertaining type. To give an example, a direct proof (p_A, p_B) of a proposition $A \wedge B$ is obtained from a direct proof p_A of A and a direct proof p_B of B. In the case of a conditional $A \supset B$ a direct proof consists of a construction f which when fed a direct proof of A returns a direct proof of B. However, the notion of a direct proof is too restrictive. Even from an intuitionistic point of view there are perfectly legitimate proofs which cannot be obtained in this direct way. For instance, in arithmetic we frequently infer a statement $F(n, m)$ for large numbers n, m from a proof of the universal statement $(\forall x{\in}\mathbb{N})(\forall y{\in}\mathbb{N})F(x, y)$ by instantiation. Or, taking an example from [38], p. 21, "we may assert even intuitionistically that $A(n) \vee B(n)$ for some numeral n without knowing a proof of $A(n)$ or $B(n)$; it would be sufficient, e.g., if we know a proof of $A(0) \vee B(0)$ and a proof of $(\forall x{\in}\mathbb{N})\big([A(x) \vee B(x)] \rightarrow [A(x+1) \vee B(x+1)]\big)$." Thus from an intuitionistic point of view it is sufficient that we know a procedure to obtain a direct proof but it is not necessary that such a direct proof be actually constructed. As a result, the problems met with (MT_1) can be remedied by saying:

(MT_2) To know the meaning of A is to know the conditions for asserting A.
The condition for asserting A is that we either know a canonical proof of A *or* know a procedure for obtaining such a proof.

4.3.4 On the Harmony Between the Rules for Making Assertions and Drawing Consequences

Dummett points out that the view expressed in (MT_2) may be somewhat immature in that it neglects an important aspect of the use of an assertive sentence which manifests itself in the commitments made by asserting it. In the case of mathematics, the full grasp of the meaning of a statement not only comprises the ability to recognize a proof of it when one is presented to us but also our capacity to draw consequences from it. He therefore suggests that in a theory of meaning a harmony should obtain between the rules for asserting a sentence and the rules for drawing consequences from it. Thus, in addition to the molecular view, there is an important strand of ideas, originating with Gentzen, running through Dummett's theory of meaning. Gentzen's elimination rules are rules for inferring consequences from a sentence and run parallel to the introduction rules. Hence, taking elimination rules into account we arrive at an enriched concept of meaning of the following form:

(MT_3) To know the meaning of A is to know the conditions for asserting A.
The condition for asserting A is that we either know a canonical proof of A *or* know a procedure for obtaining such a proof, and, moreover, that we know the rules for drawing consequences from A.

(M_3) doesn't specify the exact nature of the rules for drawing consequences from A. In a sense, there are many ways of inferring further sentences from a given one. Nonetheless, the elimination rules pertaining to A seem to occupy a privileged position in that they flow naturally from the corresponding introduction rules. Such a view seems to prevail in **MLTT** where this idea is developed at full scale. Prawitz embraces the notion of meaning as developed in (M_3) as the most desirable candidate but also enters the caveat that "in order for such a meaning theory to appear reasonable, it seems that one has to argue for the unique position of the elimination rules." ([38], p. 33) He conjectures that what distinguishes the elimination rules from other possible candidates might be that they constitute in some sense the strongest rules for drawing consequences from an assertion.

To my mind the question of the position of the elimination rules is best discussed in a wider context where one views propositions as types and in addition to the type forming operations pertaining to the BHK-interpretation takes also mathematically important types into account. These so-called *inductive data types* will be discussed in the next subsection. In the case of inductive data types, the elimination rules amount to the central tool for performing mathematical operations on them. They enable one to prove assertions about all the elements of the type by structural induction and at the same time allow one to define functions on the elements of the type by structural recursion. Owing to in-depths proof-theoretic research it is known that the omission of elimination rules in formal systems featuring inductive types gives rise to theories of weak proof-theoretic strength whereas the inclusion of elimination rules results in theories of considerable proof-theoretic strength. The latter clearly corroborates the view that elimination rules are pivotal in determining the meaning of statements involving mathematical types as e.g. the natural numbers.

4.4 Inductive Data Types

Thus far I have only addressed a small fragment of **MLTT** which is sufficient for developing logic. The concept of an *inductive data type* is central to Martin-Löf's constructivism. It was very succinctly stated by Gödel in a handwritten text for an invited lecture which he delivered in 1933. Therein he described constructive mathematics by the following two characteristics:

1. *The application of the notion "all" or "any" is to be restricted to those infinite totalities for which we can give a* **finite procedure for generating all their elements** *(as we can, e.g., for the totality of integers by the process of forming the next greater integer and as we cannot, e.g., for the totality of all properties of integers).*

2. *Negation must not be applied to propositions stating that something holds for all elements, because this would give existence propositions. Or to be more exact: Negatives of general propositions (i.e., existence propositions) are to have meaning in our system only in the sense that we have found an example but, for the sake of brevity, do not state it explicitly. I.e., they serve merely as an abbreviation and could be entirely dispensed with if we wished.*
From the fact that we have discarded the notion of existence and the logical rules concerning it, it follows that we are left with essentially only one method for proving general propositions, namely, **complete induction applied to the generating process of our elements.** [...] *and so we may say that our system is based exclusively on the* **method of complete induction in its definitions as well as its proofs.** *([22], p. 51)*

The paradigm of an inductive data type is that of the type of natural numbers. The elimination rules for $\mathbb{N}$ correspond to the principles of induction and recursion over $\mathbb{N}$. Likewise, for an arbitrary inductive type A the elimination rules encapsulate the principles of structural induction and recursion over A. The induction principle for A tells one how to prove properties for all elements of a type by induction on the build-up of its canonical elements and, correspondingly, the recursion principle tells one how to define a function on all canonical elements of the type by recursion on their build-up. More formally, elimination rules explain how non-canonical elements are formed via the selector, and the equation rules explain how the selector operates on the canonical elements. The selector pertaining to the type $\mathbb{N}$ is $\mathbf{R}_{\mathbb{N}}$. The elimination rule is the following:

$$(\mathbb{N}\text{-elimination}) \quad \frac{c : \mathbb{N} \qquad d : C(0) \qquad \begin{matrix}[x, \mathbb{N}, y : C(x)] \\ e(x, y) : C(Sx)\end{matrix}}{\mathbf{R}_{\mathbb{N}}(c, d, (x, y))e(x, y)) : C(c)}$$

$\mathbf{R}_{\mathbb{N}}(c, d, (x, y)e(x, y))$ is explained as follows: first execute c, getting a canonical element of $\mathbb{N}$, which is either 0 or Sa for some $a : \mathbb{N}$. In the first case, continue by executing d, which yields a canonical element $f : C(0)$; but, since $c = 0 : \mathbb{N}$ in this case, f is also a canonical element of $C(c) = C(0)$. In the second case, substitute a for x and $\mathbf{R}_{\mathbb{N}}(a, d, (x, y)e(x, y))$ (namely, the preceding value) for y in $e(x, y)$ so as to get $e(a, \mathbf{R}_{\mathbb{N}}(a, d, (x, y)e(x, y))$. Executing it, we get a canonical f which, by the right premiss, is in $C(Sa)$ by the first case. Otherwise, continue as in the second case, until we eventually reach the value 0. This explanation of the above elimination rule also explains the equality rules.

$$(\mathbb{N}\text{-equality I}) \quad \frac{d : C(0) \qquad \begin{matrix}[x : A, y : C(x)] \\ e(x, y) : C(Sx)\end{matrix}}{\mathbf{R}_{\mathbb{N}}(c, d, (x, y)e(x, y)) = d : C(c)}$$

$$(\mathbb{N}\text{-equality II}) \quad \frac{a : \mathbb{N} \qquad d : C(0) \qquad \begin{matrix}[x : \mathbb{N}, y : C(x)] \\ e(x, y) : C(Sx)\end{matrix}}{\mathbf{R}_{\mathbb{N}}(sa, d, (x, y)e(x, y)) = e(a, \mathbf{R}_{\mathbb{N}}(a, d, (x, y)e(x, y))) : C(Sa)}$$

An infinitary example of an inductive definition in Martin-Löf type theory is the W-type built over a family of types. A special case of the latter is the type of Brouwer ordinals or well-founded trees over **N** which may be taken as a constructive version of the second number class of ordinals. The inductive generation proceeds by the following introduction rules:

$$\bar{0} : \mathcal{O} \qquad \frac{a : \mathcal{O}}{a' : \mathcal{O}} \qquad \frac{f : \mathbb{N} \to \mathcal{O}}{\sup(f) : \mathcal{O}}$$

An inductive definition often involves the characterization of a collection of objects as the smallest collection satisfying certain closure conditions. In classical set theory the inductively defined set is usually obtained as the intersection of all collections that satisfy the closure conditions. Such an explicit definition is thoroughly impredicative in that the collection is defined using quantification over all collections. It would seem therefore that if one wants to make sense of infinitary inductive definitions it is necessary at some point to make use of some impredicative definitions. This is certainly the way that inductive definitions are accounted for in classical set theory. But there is something unsatisfying about this conclusion. Many inductive definitions can be intuitively understood directly in their own terms and impredicative definitions are only required in order to represent them within a particular framework such as classical set theory.

4.5 Reflection via Universes

The openendedness of Martin-Löf type theory is particularly manifest in the introduction of so-called *universes*. Type universes encapsulate the informal notion of reflection whose role may be explained as follows. During the course of developing a particular formalization of type theory, the type theorist may look back over the rules for types, say $\mathcal{C}$, which have been introduced hitherto and perform the step of recognizing that they are valid according to Martin-Löf's informal semantics of meaning explanation. This act of 'introspection' is an attempt to become aware of the conceptions which have governed our constructions in the past. It gives rise to a *"* **reflection principle** *which roughly speaking says whatever we are used to doing with types can be done inside a universe"* ([29], p. 83). On the formal level, this leads to an extension of the existing formalization of type theory in that the type forming capacities of $\mathcal{C}$ become enshrined in a type universe $\mathbf{U}_{\mathcal{C}}$ mirroring $\mathcal{C}$. If, e.g., $\mathcal{C}$ consists of the type forming operations $\mathbb{N}, +, \Pi$ and their rules, then this gives rise to the rules:

$$(\mathbf{U}_{\mathcal{C}}\text{-formation}) \qquad \mathbf{U}_{\mathcal{C}} : type \qquad \frac{a : \mathbf{U}_{\mathcal{C}}}{\mathbf{T}_{\mathcal{C}}(a) : type}$$

$$(\mathbf{U}_{\mathcal{C}}\text{-introduction}) \qquad \hat{\mathbb{N}} : \mathbf{U}_{\mathcal{C}} \qquad \mathbf{T}_{\mathcal{C}}(\hat{\mathbb{N}}) = \mathbb{N}$$

$$\frac{a : \mathbf{U}_{\mathcal{C}} \quad b : \mathbf{U}_{\mathcal{C}}}{a \hat{+} b : \mathbf{U}_{\mathcal{C}}} \qquad \frac{a : \mathbf{U}_{\mathcal{C}} \quad b : \mathbf{U}_{\mathcal{C}}}{\mathbf{T}_{\mathcal{C}}(a \hat{+} b) = \mathbf{T}_{\mathcal{C}}(a) + \mathbf{T}_{\mathcal{C}}(b)}$$

$$\frac{a : \mathbf{U}_{\mathcal{C}} \qquad \begin{array}{c}[x : \mathbf{T}_{\mathcal{C}}(a)] \\ t(x) : \mathbf{U}_{\mathcal{C}}\end{array}}{\hat{\Pi}(a, (\lambda x)t(x)) : \mathbf{U}_{\mathcal{C}}} \qquad \frac{a : \mathbf{U}_{\mathcal{C}} \qquad \begin{array}{c}[x : \mathbf{T}_{\mathcal{C}}(a)] \\ t(x) : \mathbf{U}_{\mathcal{C}}\end{array}}{\mathbf{T}_{\mathcal{C}}(\hat{\Pi}(a, (\lambda x)t(x))) = (\Pi x : \mathbf{T}_{\mathcal{C}}(a))\mathbf{T}_{\mathcal{C}}(t(x))}$$

$x : \mathbf{U}_{\mathcal{C}}$ and $\mathbf{T}_{\mathcal{C}}(x)$ are defined by a simultaneous induction. The elements of type $\mathbf{U}_{\mathcal{C}}$ are codes or Gödel numbers of types generated from $\mathbb{N}$ via the type forming operation $+$ and Π. Each time an element of $\mathbf{U}_{\mathcal{C}}$ is generated there is a declaration, by means of the decoding function $\mathbf{T}_{\mathcal{C}}$, of the type for which it stands.

It might seem that type universes are not that different from inductive data types. However, as opposed to inductive data types their elements are not simply defined by a positive inductive definition. They incorporate the extra feature of being equipped with a type-valued function defined on them, and, moreover, this simultaneously defined decoding function is allowed to occur negatively in its introduction rules as, e.g., in those for $\hat{\Pi}$.

As in the case of inductive data types, the introduction rules for type universes give rise to their dual, the elimination rules, which entail that the elements of $\mathbf{U}_{\mathcal{C}}$ are exactly generated by the forms of type that $\mathbf{U}_{\mathcal{C}}$ is reflecting.[11]

4.6 Reflection in Higher Order Universes

It is the nature of reflection to aim at higher degrees of introspection. The first level is encapsulated in universes containing codes for types which reflect previously accepted type forming operations. In [29, 31] Martin-Löf considered an infinite, externally indexed tower of universes $\mathbf{U}_0 \in \mathbf{U}_1 \in \cdots \in \mathbf{U}_n \in \ldots$ all of which are closed under the same standard ensemble of set forming operations. The next natural step was to implement a *universe operator* into type theory which takes a family of sets and constructs a universe above it. Such a universe operator was formalized by Palmgren while working on a domain-theoretic interpretation of the logical framework with an infinite sequence of universes (cf. [35]). Aiming at extensions of type theory with more powerful axioms, Martin-Löf then suggested finding axioms for a universe $\mathbb{V}$ which itself is closed under the universe operator. The type-theoretic formalization of the pertinent rules is due to Palmgren [36], where the universe was referred to as a *superuniverse* for intuitionistic type theory.

[11] Martin-Löf has expressed the view that it is natural to keep the type $\mathbf{U}_{\mathcal{C}}$ open to reflect any additional forms of type that can arise in the future and therefore not to incorporate elimination rules for universes. In my view such a decision is a mere convenience lest one be forced to specify different forms of universes.

After the superuniverse one can go to a new level of abstraction which consists in gaining insight into how type forming operations are obtained. One can, e.g., internalize the introduction of new universe operators in a formal system of type theory wherein each operator $F : \mathbf{Fam} \to \mathbf{Fam}$ from families of types to families of types gives rise to a new universe operator $\mathbb{S}(F) : \mathbf{Fam} \to \mathbf{Fam}$. When applied to a family of sets, $\mathbb{S}(F)$ produces a universe above it which, in addition to the standard set constructors, is closed under F. Such a system, denoted **MLF**, was presented in [45].

A further level of abstraction gives rise to types having codes for such operations as elements which may be called universes of type two. Such a step has, e.g., been taken in [43] which introduces a formal type theory **MLQ**. Central to **MLQ** are a universe **M** and a universe of codes for type constructors **Q** which are defined simultaneously. In more generality, this idea has been developed by Palmgren (cf. [36]) who introduced type universes of arbitrary finite levels.

Essentially stronger universe constructions require new ideas and their potential seems to be limited by creativity only. However, in [45] I have stated a conjecture about the proof-theoretic strength of universe constructions based on higher types which predicts that they can always be mimicked in a classical set theory called **KPM**. **KPM** was designed to formalize segments L_μ of the constructible hierarchy L, where μ is a recursively Mahlo ordinal (cf. [39]).

Another important step in advancing constructive type theory has been taken by Setzer who gave a formalization of a Mahlo universe in type theory (cf. [51]), dubbed **TTM**. Constructive Zermelo-Fraenkel set theory conjoined with an axiom asserting the existence of a Mahlo set can be interpreted in **TTM** as has been shown in [46]. Setzer's theory is stronger than those based on higher type universes. It provides an important step for expanding the realm of Martin-Löf type theory. The difference between **TTM** and the systems above is that **TTM** introduces a new construction principle which is not foreshadowed in Martin-Löf's original papers. This is witnessed by the fact that models for Setzer's Mahlo universe are generated by a non-monotonic inductive definition (see Section 6) and, furthermore, by an observation due to Palmgren which shows it to be incompatible with elimination rules for the universe. In a sense, **TTM** means a paradigm shift to a *new* Martin-Löf type theory in that the rules for forming the elements of a type are no longer required to be monotonic. Palmgren's proof that the usual elimination rules for universes yield an inconsistency when applied to **TTM** prompted Martin-Löf to respond that universes need not be equipped with elimination rules and that such rules may be rather alien to the idea of a universe. From the point of view of the classical theory of inductive definitions (cf. [25]) though, Palmgren's inconsistency proof presents no surprise. The usual elimination rules for universes are tailored for monotone inductive definitions and thus spring from the crucial property that the inductively defined type (or set) is the least fixed point of the corresponding operator. To be more precise, in set theory a *monotone inductive definition* over a given set A is derived from a mapping

$$\Phi : \mathbf{pow}(A) \to \mathbf{pow}(A)$$

that is monotone, i.e., $\Phi(X) \subseteq \Phi(Y)$ whenever $X \subseteq Y \subseteq A$. Here $\mathbf{pow}(A)$ denotes the class of all subsets of A. The set inductively defined by Φ, Φ^∞, is the smallest set Z such that $\Phi(Z) \subseteq Z$. Due to the monotonicity of Φ such a set exists. In the case of Setzer's type **TTM**, however, the pertinent operator is non-monotone and does not possess a least fixed point, hence the inconsistency arises. The classical view on non-monotone inductive definitions is that the inductively defined set is obtained in stages by iteratively applying the corresponding operator to what has been generated at previous stages along the ordinals until no new objects are generated in this way. More precisely, if $\Phi : \mathbf{pow}(A) \to \mathbf{pow}(A)$ is an arbitrary mapping then the set-theoretic definition of the set inductively defined by Φ is given by

$$\Phi^\infty := \bigcup_\alpha \Phi^\alpha,$$

$$\Phi^\alpha := \Phi\Big(\bigcup_{\beta<\alpha} \Phi^\beta\Big) \cup \bigcup_{\beta<\alpha} \Phi^\beta,$$

where α ranges over the ordinals. I agree with Setzer (cf. [51], p. 157) that the correct type-theoretic elimination rules for non-monotone type universes have yet to be found. Furthermore, I'd like to conjecture that the "correct" elimination rules can be found by viewing non-monotone inductive types as being equipped with a W-type (or well-ordering) which provides the stages of its inductive generation.

5 Which Fragments of Second Order Arithmetic are Secured by MLTT?

The central notions of *proof-theoretic reducibility* and *proof-theoretic strength* will be used below. For the readers convenience I shall insert a brief account of them. All theories T considered in the following are assumed to contain a modicum of arithmetic. For definiteness let this mean that the system **PRA** of Primitive Recursive Arithmetic is contained in T, either directly or by translation.

Definition 1 Let T_1, T_2 be a pair of theories with languages $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, and let Φ be a (primitive recursive) collection of formulae common to both languages. Furthermore, Φ should contain the closed equations of the language of **PRA**.

We then say that T_1 is *proof-theoretically* Φ*-reducible to* T_2, written $T_1 \leqslant_\Phi T_2$, if there exists a primitive recursive function f such that

$$\mathbf{PRA} \vdash \forall\phi \in \Phi\, \forall x\, [\mathrm{Proof}_{T_1}(x, \phi) \;\to\; \mathrm{Proof}_{T_2}(f(x), \phi)]. \tag{1}$$

T_1 and T_2 are said to be *proof-theoretically* Φ*-equivalent*, written $T_1 \equiv_\Phi T_2$, if $T_1 \leqslant_\Phi T_2$ and $T_2 \leqslant_\Phi T_1$.

The appropriate class Φ is revealed in the process of reduction itself, so that in the statement of theorems we simply say that T_1 is *proof-theoretically reducible* to T_2 (written $T_1 \leqslant T_2$) and T_1 and T_2 are *proof-theoretically equivalent* (written $T_1 \equiv T_2$), respectively. Alternatively, we shall say that T_1 and T_2 have the *same proof-theoretic strength* when $T_1 \equiv T_2$.

Proof-theoretic investigations undertaken by Buchholz, Pohlers and Sieg in the late 1970s (see [8]) showed that the intuitionistic theory of below ϵ_0-iterated inductive tree classes, $\mathbf{ID}^i_{<\epsilon_0}(\mathcal{O})$, is of the same proof-theoretic strength as (Σ^1_2-**AC**). With the help of one type universe reflecting the W-type one can interpret $\mathbf{ID}^i_{<\epsilon_0}(\mathcal{O})$ in type theory thereby reducing (Σ^1_2-**AC**) to type theory. However, a considerably stronger theory than (Σ^1_2-**AC**) can be proven consistent in **MLTT**. Jäger and Pohlers [27]) gave an ordinal analysis (Σ^1_2-**AC**) plus bar induction in terms of an ordinal representation system. Utilizing the pertaining ordinal representation system the following result was obtained independently by Rathjen [40] and Setzer [49]:

Theorem 1 *The consistency of* (Σ^1_2-**AC**)+**TI** *is provable in Martin-Löf's 1984 type theory.*

On the part of the intuitionists/constructivists, the following objection could be raised against the significance of consistency proofs: even if it had been constructively demonstrated that the classical theory T cannot lead to mutually contradictory results, the theorems of T would nevertheless be propositions without sense and their investigation therefore an idle pastime.[12] Well, it turns out that the constructive well-ordering proof of the representation system used in the analysis of (Σ^1_2-**AC**)+**TI** yields more than the mere consistency of the latter system. For the important class of Π^0_2 statements one obtains a conservativity result.

Theorem 2 (Rathjen [40]; Setzer [50])

- *The soundness of the negative arithmetic fragment of* (Σ^1_2-**AC**)+**TI** *is provable in Martin-Löf's 1984 type theory.*
- *Every* Π^0_2 *statement provable in* (Σ^1_2-**AC**)+**TI** *has a proof in Martin-Löf's 1984 type theory.*

Theorem 1 is by no means the strongest result of its kind. The paper [42] introduces a theory of second order arithmetic which is based on (Σ^1_2-**AC**)$\restriction$ but in addition has axioms stating that there exist many β-models of (Σ^1_2-**AC**). A β-model is a model with respect to which the notion of well-foundedness is absolute. More precisely,

[12] This is reminiscent of Russell's bon mot that not everything in an inconsistent theory can be true but that every axiom of a consistent theory may be false.

let **T** be the theory (Σ^1_2-**AC**) $\restriction$ plus the scheme asserting that every true Π^1_3 statement is reflected in a β-model of (Σ^1_2-**AC**). It follows from [42], Theorem 5.15 that the type theory **MLF** and **T** have the same proof-theoretic strength. It also follows from [42], Theorem 5.15 that the type theory **MLQ** proves the consistency and Π^0_2 soundness of T.

6 The Limits of Martin-Löf Type Theory

Above we have seen that ordinary mathematics is demonstrably consistent relative to **MLTT**. Thus from the point of view of justifying mathematical practice in a Hilbertian way the existing formalizations of **MLTT** are already powerful enough. Notwithstanding it is still of great interest to ponder where the limits of **MLTT** lie. This is a somewhat ambiguous question, though. It is, of course, fundamental to Martin-Löf's outlook that it is possible to apply reflection to any particular formalization of **MLTT**, and thus obtain a stronger system. Consequently, Martin-Löf type theory as a foundational undertaking cannot be captured by a formal system that the type theorist can recognize as being sound. Moreover, it doesn't seem to be possible to single out a particular fragment T_M of **ZFC** and argue that **MLTT** exhausts and furthermore is equiconsistent with T_M since drawing such an exact limit to **MLTT** appears to entail that this limit could be approached from within Martin-Löf type theory and therefore be overcome.

Nonetheless, it appears to be possible to transcend the type theorists realm by adopting, as it were, a hypothetical 'eagles' point of view as in classical set theory and reflect from such an advanced position on all possible moves the type theorist can ever perform. This thought experiment should then allow one to delineate bounds for the realm which the type theorist will never be leaving, or, more precisely, it should enable the Cantorian set theorist to draw a line in his set-theoretic world beyond which the type theorist will never be able to reach. An intellectual difficulty in pursuing this project is to become aware of the conceptions which govern all possible thinking of a Martin-Löf type theorist. A first and rather coarse reflection suggests three basic principles of Martin-Löf type theory.

- (A0) (Predicativism) The realm of types is built in stages (by the idealized type theorist). It is not a completed totality. In declaring what are the elements of a particular type it is disallowed to make reference to all types.
- (A1) A type A is defined by describing how a canonical element of A is formed as well as the conditions under which two canonical elements of A are equal.
- (A2) The canonical elements of a type must be namable, that is to say, they must allow for a symbolic representation, as a word in a language whose alphabet, in addition to countably many basic symbols, consists of the elements of previously introduced types. Here "previously" refers back to the stages of (A0).

Principles (A0), (A1), and (A2) are liable to different readings and are in need of further rumination. The predicativism of **MLTT** is much more liberal than the one based on autonomous progressions of theories delineated in Feferman's and Schütte's work. In addition to the inductively defined set of natural numbers it allows for other inductive definitions of a rather general kind. This type of predicativism is often referred to as "generalized predicative".

The types of **MLTT** are comprised roughly of three different kinds: explicitly defined types (e.g. the empty type and the type of Booleans N_1) as well types defined explicitly from given types or families of types (e.g. $A + B$, $(\Sigma x : A)B(x)$), inductively defined types, and functions types (e.g. $A \to B$, $(\Pi x : A)B(x)$). The understanding of function types in **MLTT** requires an elucidation of the notion of function. In [32] the notion of function is explained in terms of the notion of hypothetical proof which is declared to be a primitive notion (see [32], p. 41). Similarly, ([31], p. 23) Martin-Löf writes that: *"The reason that B^A can be constructed as a set is that we take the notion of function as primitive, instead of defining a function as a set of ordered pairs or a binary relation satisfying the usual existence and uniqueness conditions, which would make it a category (like $\mathcal{P}(A)$) instead of a set."* These explanations, however, are not very informative and the latter seems to be rather misguided as the distinction between a set of functions $\{0, 1\}^{\mathbb{N}}$ and a proper class of subsets of $\mathbb{N}$ can be made in constructive set theories notwithstanding that functions are defined as sets of ordered pairs (see [1]).

The axiom that the class of functions $A \to B$ of a set A into a set B is a set is known as the *Exponentiation* axiom. The difference between the axioms of Power set and Exponentiation was explained by Myhill ([33], p. 351) as follows: *"Power set seems especially nonconstructive and impredicative compared with the other axioms: it does not involve, as the others do, putting together or taking apart sets that one has already constructed but rather selecting, out of the totality of all sets those that stand in the relation of inclusion to a given set. One could make the same, admittedly vague, objection to the existence of the set $A \to B$ of mappings of A to B but I do not think the situation is parallel–a mapping or function is a rule, a finite object which can actually be given;."*

To be able draw any limits of **MLTT** within classical set theory we must be able to understand how functions are conceived in **MLTT**. If the function type $A \to B$ is to be taken in the classical set-theoretic way it will be a fully impredicative set and the difference between **MLTT** and classical set theory wanes in importance. The charge of impredicativity against the intuitionistic notion of proof is well known. The problem is vividly related in ([12], pp. 274–275): *"This may make it appear that, in order to recognize an operation as a proof of a statement of the form $(B \to C) \to D$, we must survey all possible proofs of $B \to C$ to which the operation might be applied; and in order to do this, we should need to know the whole of (existing) mathematics, since we cannot tell on what mathematical results a recognition of the efficacy of an operation proving $B \to C$ might not draw. If this were so, the intuitionistic acceptance of compositionality and consequent rejection of mathematical holism would be spurious: an understanding of $(B \to C) \to D$ would not rest on a* prior *understanding of $B \to C$, since an understanding of*

$B \to C$ would already involve a knowledge of all mathematics. Futhermore, meaning and even proof itself would be unstable. As mathematics advances, we become able to conceive of new operations and to recognize them and others as effectively transforming proofs of B into proofs of C: and so the meaning of $B \to C$ would change, if a grasp of it required us to circumscribe such operations in thought. Moreover, an operation which would transform proofs of $B \to C$ available to us now into a proof of D might not so transform proofs of $B \to C$ which became available to us with the advance of mathematics: and so what would now count as a valid proof of $(B \to C) \to D$ would no longer count as one." Dummett ([12], p. 274) assures us that these fears are groundless: *"In order to recognize an operation as a proof of $(B \to C) \to D$, we must think of it as acting on anything we may ever recognize as a proof of $B \to C$. Of such a proof, we know in advance only what is specified by the intuitive explanation of $\to$: namely, that we recognize it as an effective operation, and as one that will transform any proof of B into a proof of C."* It's not obvious to me what to make of Dummett's rebuttal of the charge of impredicativity but one aspect about it seems to be clear enough, namely that in intuitionistic mathematics operations must be effective. This is also confirmed on the next page [12], p. 275: *"An operation f defined over a domain D carries each element x of D into an element $f(x)$ of its range R. In intuitionistic mathematics the operation must be given as an effective means of determining the result $f(x)$ from the way in which x is given ... ."*

[34] (p. 48) offers the following explanation of the concept of function in **MLTT**: *"The basic notion of function is an expression formed by abstraction."* Martin-Löf's remark about function types (or sets) in [31] (p. 27) is somewhat cryptic and seems to support an impredicative reading: *"Since, in general, there are no exhaustive for generating all functions from one set to another, it follows that we cannot generate inductively all the elements of a set of the form $(\Pi x \in A)B(x)$ (or, in particular, of the form B^A, like N^N)."* However, this quote allows for a much more benign interpretation. In **MLTT** the canonical objects of $A \to B$ are formed by abstraction $(\lambda x)t(x)$ from functional expressions (terms) $t(x)$. As Martin-Löf type theory is an open-ended system, the addition of new types with their pertaining constructors and selectors is always possible and will give rise to new terms involving these constants. As a result, terms will be created which, by putting a lambda in front of them, may give rise to new canonical elements of function types $A \to B$ created at earlier stages. This seeming circularity in defining elements of a type after the creation of this type not only affects function types but also inductively defined types like the W-types. To my mind, the distinction between canonical and non-canonical elements given by Martin-Löf is not a happy one in the case of function types. The meaning of an implication $A \supset B$ is explained by saying what counts as a canonical proof of it. If the meaning depended on all that counts as a proof of it, then the meaning would change each time we found a new proof of it. However, in **MLTT** any non-canonical proof b of the proposition $A \supset B$ gives rise to the canonical proof $(\lambda x)Ap(b, x)$ of $A \supset B$, rendering the distinction between canonical and non-canonical proofs just a formal one. Be that as it may, there is agreement among intuitionists and constructivists that a function f is always given by rule

which effectively determines the result $f(x)$ for every argument x in the domain of f. In **MLTT** functions are presented by functional expressions that encapsulate a programme for calculating $f(x)$ for x in the domain of f. In light of this, the main argument for conceiving of function types as predicative will be based on the tenet that functions be concrete objects given by rules. At first blush, this appears to entail that all functions from $\mathbb{N}$ to $\mathbb{N}$ must be recursive, and thus point to Church's thesis. There are, however, cherished principles of Brouwerian mathematics such as the Fan theorem that are incompatible with CT. Also the extensional version of Martin-Löf type theory refutes CT for the simple reason that the propositions-as-types interpretation forces CT to be a $\Pi\Sigma$-type whose inhabitedness conjoined with the principle of extensionality for functions with domain $\mathbb{N} \to \mathbb{N}$ implies the solvability of the halting problem. This result, though, does not mean that extensional **MLTT** proves the existence of non-recursive functions as one can easily construct models of **MLTT** with extensionality wherein all functions are recursive. Furthermore, intensional **MLTT** is compatible with CT.

The problem thus remains to delineate a class of functions that comprises all functions acceptable in Martin-Löf type theory. I will argue that all functions that deserve to be called effective must at least be definable in a way that is persistent with expansions of the universe of types.

To put flesh to this idea, I consider it fruitful to investigate a rigorous model of the principles underlying **MLTT** within set theory. In the following, let us adopt a classical Cantorian point of view and analyze the principles (A0),(A1),(A2) on this basis. Firstly, types are to be interpreted as sets. By Gödel numbering, (A2) hereditarily has the consequence that nothing will be lost by considering all types to be surjective images of subsets of $\mathbb{N}$. In combination with (A1), such an encoding yields that every inductive type A can be emulated by an inductive definition Φ over the natural numbers together with a decoding function D, where

$$\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$$

is a (class) function from the class of all subsets of $\mathbb{N}$, $\mathbf{pow}(\mathbb{N})$, to $\mathbf{pow}(\mathbb{N})$. The set inductively defined by Φ, Φ^∞, has the set-theoretic definition

$$\begin{aligned} \Phi^\infty &:= \bigcup_\alpha \Phi^\alpha, \\ \Phi^\alpha &:= \Phi\Big(\bigcup_{\beta<\alpha} \Phi^\beta\Big) \cup \bigcup_{\beta<\alpha} \Phi^\beta, \end{aligned}$$

where α ranges over the ordinals. The elements of type A are then considered to be represented by the elements of Φ^∞. Thus the type A will be identified with the set

$$\{D(x) \,:\, x \in \Phi^\infty\}.$$

The vast majority of inductively defined types of **MLTT** is given by monotone operators[13] Φ whose iterations satisfy $\Phi^\alpha = \Phi\big(\bigcup_{\beta<\alpha} \Phi^\beta\big)$. But since Setzer's type theory **TTM** (cf. [51]) features a type which is not generated by a monotone operator, I shall not impose the restriction of monotonicity on operators.

A further step in delineating **MLTT** consists in describing the allowable operators Φ and decoding functions D_A. A common way of classifying inductive definitions proceeds by their syntactic complexity. To find such a syntactic bound it is in order to recall that the type theorists develop their universe of types in stages. Introducing a new type A consists in describing a method for generating its elements. Taking into account that the type-theoretic universe is always in a state of expansion it becomes clear that each time a new element of A is formed by the method of generation for A, this method can only refer to types that have been built up hitherto. Furthermore, the method of generation of elements should also obey a persistency condition of the following form: If at a certain stage an object t is recognized as an element of A then an expansion of the type-theoretic universe should not nullify this fact, i.e. the method should remain to be applicable and yield $t : A$ in the expanded universe as well. And in the same vein, if A is a type of codes of types which comes endowed with a type-valued decoding function D (like in the case of type universes), then the validity of equations between types of the form $D(x) = B$ with $x : A$ should remain true under expansions of the universe of types.

Framing the foregoing in set-theoretic terms amounts to saying that the truth of formulas describing $t \in \Phi(X)$ and $D(t) = b$, respectively, ought to be persistent under adding more sets to a set-theoretic universe. In more technical language this means that whenever $\mathbb{M}$ and $\mathbb{P}$ are transitive sets of sets such that $t, X \in \mathbb{M}$, $\mathbb{M} \subseteq \mathbb{P}$ and $(\mathbb{M}, \in_{\mathbb{M}}) \models t \in \Phi(X)$, then $(\mathbb{P}, \in_{\mathbb{P}}) \models t \in \Phi(X)$[14] should obtain as well. The same persistency property should hold for formulas of the form '$D(x) = b$'.

The formulas which can be characterized by the latter property are known in set theory as the Σ-formulas. They are exactly the collection of set-theoretic formulas generated from the atomic and negated atomic formulas by closing off under $\wedge$, $\vee$, bounded quantifiers $(\forall x{\in}a)$, $(\exists x{\in}a)$ and unbounded existential quantification $\exists x$ (cf. [2], I.8).

In view of the preceding, one thus is led to impose restrictions on the complexity of inductive definitions for generating types in **MLTT** as follows.

(A3) Every inductive definition $\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$ for generating the elements of an inductive type A in **MLTT** and its pertinent decoding function are definable by set-theoretic Σ-formulas. These formulas may contain further sets as parameters, corresponding to previously defined types.

To avoid misunderstandings, I'd like to emphasize that (A3) is not meant to say that every such Σ inductive definition gives rise to a type acceptable in **MLTT**. (A3) is intended only as a delineation of an upper bound.

---

[13] Φ is said to be *monotone* if $X \subseteq Y$ implies $\Phi(X) \subseteq \Phi(Y)$.

[14] $\in_{\mathbb{M}}$ stands for the elementhood relation restricted to sets in $\mathbb{M}$.

Having determined that the case for the predicativity of function types rests on the requirement that functions be given by rules that enable one to compute their values effectively, it is plain that any such function must be definable in an absolute way. In view of the foregoing arguments for restrictions imposed on inductive types in conjunction with (A2) one is led to require the following:

(A4) All the functions figuring in **MLTT** belong to the set

$$\mathbf{Func} := \{f \subseteq \mathbb{N} \times \mathbb{N} : f \text{ is a } \Sigma\text{-definable function}\}.$$

Note that the functions in **Func** are required to have a lightface Σ definition, that is to say definitions must not involve parameters (oracles).

The functions in **Func** are known from generalized recursion theory on ordinals. **Func** consists all ∞-partial recursive functions from $\mathbb{N}$ to $\mathbb{N}$ (see [25]). In terms of the analytical hierarchy, **Func** can be characterized as the class of all (lightface) Σ^1_2-definable partial functions from $\mathbb{N}$ to $\mathbb{N}$.

(A4) and previous considerations induce us to delineate the interpretation of product types as follows:

(A5) Every product type $(\Pi x : A)B(x)$ in **ML** is a set of functions from A to $\bigcup_{x:A} B(x)$ Σ-definable (with parameters) from previously defined types and the set **Func**. Moreover, $(\Pi x : A)B(x)$ is a subset of **Func**.

The principles (A0)–(A5) will allow us to draw a limit to **ML** in the guise of a small fragment of **ZF**. This fragment, notated **T**, will be based on the ubiquitous Kripke-Platek set theory, **KP**. Kripke-Platek set theory is a truly remarkable subsystem of **ZF**. Though considerably weaker than **ZF**, a great deal of set theory requires only the axioms of this subsystem. **KP** arises from **ZF** by omitting the power set axiom and restricting separation and collection to bounded formulas, that is formulas without unbounded quantifiers. **KP** has been a major site of interaction between many branches of logic (for more information see the book by Barwise [2]). The transitive models of **KP** are called *admissible sets*.

To describe **T**, we have to alter **KP** slightly. Among the axioms of **KP** is the foundation scheme which says that every non-empty definable class has an $\in$-least element. Let $\mathbf{KP}^r$ result from **KP** by restricting the foundation scheme to sets. In addition to $\mathbf{KP}^r$, **T** has an axiom asserting that every set is contained in a transitive set which is a Σ_1 elementary substructure of the set-theoretic universe V (written $M \prec_1 V$). To be more precise, let $M \prec_1 V$ stand for the scheme

$$\forall a{\in}M \left[\exists x\, \phi(x, a) \rightarrow \exists x \in M\, \phi(x, a)\right]$$

for all bounded formulas $\phi(x, y)$ with all free variables exhibited. Using a Σ_1 satisfaction predicate, $M \prec_1 V$ can actually be expressed via a single formula.

We take $\mathbf{T}$ to be the theory

$$\mathbf{KP}^r + \forall x\, \exists M\, \big(x \in M \ \wedge\ M \prec_1 V\big).$$

The following theorems are provable in $\mathbf{T}$.

Theorem 3 **Func** *is a set.*

Proof Let X be a set satisfying $X \prec_1 V$. One easily checks that every element of **Func** is in X since it is Σ definable without parameters. Moreover, **Func** is subset of M which is definable in M, whence **Func** is a set by bounded separation.

Theorem 4 *If* $\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$ *is definable by a* Σ *formula with parameters in* M *and* $M \prec_1 V$, *then* $\Phi^\infty \in M$.

The above theorem supports the claim that everything a Martin-Löf type theorist can ever develop can be emulated in $\mathbf{T}$ or, to put it more pictorially, that the boundaries of the type theorist world are to be drawn inside M, where M satisfies $M \prec_1 V$.

Before elaborating further on this question, it might be interesting to give an equivalent characterization of $\mathbf{T}$ which is couched in terms of subsystems of second order arithmetic.

Theorem 5 ([44], Theorem 3) *The theories* $(\Pi^1_2\text{-}\mathbf{CA})\restriction$ *and* $\mathbf{T}$ *prove the same statements of second order arithmetic.*

Resuming the question of the type theorist's limit, I shall now argue on the basis of $\mathbf{T}$ that every set $M \prec_1 V$ with $\mathbf{Func} \in M$ is a model of **MLTT**, i.e. it contains all the types that may ever be constructed in **MLTT**. The argument may run in this way: Types are interpreted as sets. At a certain stage the idealized type theorist, called ITT, has a certain repertoire of type forming operations, say $\mathcal{C}$. The operations correspond to a collection $\mathcal{C}_{Set}$ of Σ-definable operations on sets. Further, assume that ITT introduces a new type A by utilizing $\mathcal{C}$. Inductively we may assume that any set M with $M \prec_1 V$ and $\mathbf{Func} \in M$ is a model of ITT's reasoning as developed up to this point. Thus any such M is closed under $\mathcal{C}_{Set}$. According to (A3), the generation of the elements of A gives rise to an operator $\Phi_M : \mathbf{pow}(\mathbb{N}) \cap M \to \mathbf{pow}(\mathbb{N}) \cap M$ and a decoding function D_M which are both Σ-definable on M whenever $M \prec_1 V$. Moreover, Φ_M and D_M are uniformly definable on all $M \prec_1 V$, that is to say, there are Σ-formulas $\psi(x, y)$ and $\delta(u, v)$ such that $\Phi_M(X) = Y$ iff $(M, \in_M) \models \psi(X, Y)$ and $D_M(u, v)$ iff $(M, \in_M) \models \delta(u, v)$ whenever $X, Y \in \mathbf{pow}(\mathbb{N}) \cap M$, $u, v \in M$, and $M \prec_1 V$. Now define

$$\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$$

by letting $\Phi(X) = \Phi_M(X)$, where $X \in M$ and $M \prec_1 V$. Φ defines a function since the Φ_M are Σ definable and for every $X \subseteq \mathbb{N}$ there exists $M \prec_1 V$ such that $X \in M$. Thus $\mathbf{T}$ proves that Φ is a Σ-definable operator, i.e.,

$$\mathbf{T} \vdash \forall X \subseteq \mathbb{N}\, \exists Y\, \Phi(X) = Y.$$

Employing Theorem 4, one can deduce that Φ^∞ is a set. Moreover, as Φ^∞ is Σ definable too, one can infer that $\Phi^\infty \in M$ and thus

$$A = \{D(u) \,:\, u \in \Phi^\infty\} = \{D_M(u) \,:\, u \in \Phi^\infty_M\} \in M$$

for every $M \prec_1 V$.

7 The Higher Infinite and New Axioms

At the end of this article it is, perhaps, in order to point out that there are parts of mathematics which are permeated by set theory and thus are not capable of a constructive consistency proof. In particular the structure of sets of reals is affected by set-theoretic axioms. Examples of results that require uncountably many iterations of the power set operation are D. Martin's theorem that all Borel games are determined and Friedman's Borel diagonalization theorem.

The set existence axioms considered in modern set theory go way beyond Zermelo-Fraenkel set theory. These so-called large cardinal axioms imply that certain infinite games played on sets of reals always possess a winning strategy. Perhaps, most notably, projective determinacy, **PD**, asserts that all games which are definable in the language of $\mathbf{Z_2}$ have a winning strategy. The relation to large cardinals is that **PD** is a consequence of the existence of infinitely many Woodin cardinals. **PD** leads to a very satisfying structure theory for projective sets of reals in that under **PD** every projective set A of reals is Lebesgue measurable, has the property of Baire, and if A is uncountable, then A has a perfect subset (see [58]).

The ASL 2000 annual meeting also saw a panel discussion devoted to the question: *Does mathematics need new axioms?* (See [17]). Two of the panelists, Harvey Friedman and John Steel, maintained that mathematics needs new axioms, i.e. principles not already provable in **ZFC**. Steel's line of argument was that the descriptive set theory emanating from large cardinals is a reason why mathematicians should adopt these large cardinal axioms. Harvey Friedman held that his most recent discovery, called *Boolean relation theory* (BT), provides strong reasons for adopting new axioms as BT has consequences for the core of mathematics which are hard to dismiss. Roughly speaking, BT is concerned with the relationship between sets and their images under multivariate functions. What is most striking about Friedman's results in BT is that they encapsulate the proof-theoretic strength of certain large cardinals. Only the future can tell whether BT is ever going to play a role in the dealings of everyday mathematics.

Acknowledgement I wish to thank John Derrick for reading an earlier version of this paper, bringing several inaccuracies to my attention, and suggesting improvements. Notwithstanding that we hold differing views on the foundations of mathematics, our discussions of the paper at Monk Fryston Hall (while indulging in a cream tea) were most enjoyable.

References

1. P. Aczel, M. Rathjen: *Notes on constructive set theory*, Technical Report 40, Institut Mittag-Leffler (The Royal Swedish Academy of Sciences, 2001). *http://www.ml.kva.se/preprints/archive2000-2001.php*
2. J. Barwise: *Admissible Sets and Structures* (Springer, Berlin 1975).
3. P. Bernays: *Hilbert, David, Encyclopedia of Philosophy*, Vol. 3 (Macmillan and Free Press, New York, 1967) 496–504.
4. E. Bishop: *Foundations of Constructive Analysis* (McGraw-Hill, New York, 1967).
5. L.E.J. Brouwer: *Weten, willen, spreken* (Dutch). Euclides 9 (1933) 177–193.
6. D.K. Brown: *Functional Analysis in Weak Subsystems of Second Order Arithmetic.* Ph.D. Thesis (Pennsylvania State University, University Park,1987).
7. D.K. Brown, S. Simpson: *Which set existence axioms are needed to prove the separable Hahn-Banach theorem?* Annals of Pure and Applied Logic 31 (1986) 123–144.
8. W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: Iterated Inductive Definitions and Subsystems of Analysis (Springer, Berlin, 1981).
9. H.B. Curry, R. Feys:*Combinatory Logic, vol. I.* (North-holland, Amsterdam, 1958)
10. R. Diestel: Graph Theory (Springer, New York-Berlin-Heidelberg, 1997).
11. M. Dummett: *The philosophical basis of intuitionistic logic.* In: H.E. Rose et al.(eds.): *Logic Colloquium '73* (North-Holland, Amsterdam, 1973) 5–40.
12. M. Dummett: *Elements of Intuitionism.* 2nd edition (Clarendon Press,Oxford,2000).
13. S. Feferman: *A Language and Axioms for Explicit Mathematics*, Lecture Notes in Math. 450 (Springer, Berlin, 1975), 87–139.
14. S. Feferman: *Constructive theories of functions and classes*. In: Boffa, M., van Dalen, D., McAloon, K. (eds.), *Logic Colloquium '78* (North-Holland, Amsterdam, 1979) 159–224.
15. S. Feferman: *Why a little bit goes a long way*. In: S. Feferman: *In the Light of Logic*(Oxford University Press, Oxford, 1998).
16. S. Feferman: *Weyl vindicated: "Das Kontinuum" 70 years later.* In: S. Feferman: *In the Light of Logic*(Oxford University Press, Oxford, 1998).
17. S. Feferman, H. Friedman, P. Maddy, J. Steel: *Does mathematics need new axioms?* Bulletin of Symbolic Logic 6 (2000) 401–446.
18. H. Friedman: personal communication to L. Harrington (1977).
19. H. Friedman, N. Robertson, P. Seymour: *The metamathematics of the graph minor theorem*, Contemporary Mathematics 65 (1987) 229–261.
20. G. Gentzen: *Untersuchungen über das logische Schliessen.* Mathematische Zeitschrift 39 (1935) 176-210, 405–431.
21. G. Gentzen: *Die Widerspruchsfreiheit der reinen Zahlentheorie,* Mathematische Annalen 112 (1936) 493–565.
22. K. Gödel: *The present situation in the foundations of mathematics.* In: *Collected Works*, vol. III (Oxford University Press, New York, 1995).
23. D. Hilbert: *Über das Unendliche.* Mathematische Annalen 95 (1926) 161–190. English translation In: J. van Heijenoort (ed.): *From Frege to Gödel. A Source Book in Mathematical Logic, 1897–1931.*(Harvard University Press, Cambridge, Mass.,1967).
24. D. Hilbert and P. Bernays: *Grundlagen der Mathematik II* (Springer, Berlin, 1938).
25. P. Hinman: *Recursion-theoretic hierarchies* (Springer, Berlin, 1978).
26. W.A. Howard: *The Formulae-as-Types Notion of Construction.* (Privately circulated notes, 1969).
27. G. Jäger and W. Pohlers: *Eine beweistheoretische Untersuchung von* Δ^1_2*–CA+ BI und verwandter Systeme*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch–Naturwissenschaftliche Klasse (1982).
28. G. Kreisel: *Ordinal logics and the characterization of informal concepts of proof.* In: *Proceedings of the 1958 International Congress of Mathematicians*, (Edinburgh, 1960) 289–299.

29. P. Martin-Löf: *An intuitionistic theory of types: predicative part*. In: H.E. Rose and J. Sheperdson (eds.): *Logic Colloquium '73* (North-Holland, Amsterdam, 1975) 73–118.
30. P. Martin-Löf: *Constructive mathematics and computer programming*. In: L.J. Cohen, J. Los, H. Pfeiffer, K.-P. Podewski:*LMPS IV* (North-Holland, Amsterdam, 1982).
31. P. Martin-Löf: *Intuitionistic Type Theory*, (Bibliopolis, Naples, 1984).
32. P. Martin-Löf: *On the meanings of the logical constants and the justifications of the logical laws*, Nordic Journal of Philosophical Logic 1 (1996) 11–60.
33. J. Myhill: *Constructive Set Theory*, J. Symbolic Logic 40 (1975) 347–382.
34. B. Nordström, K. Petersson and J.M. Smith: *Programming in Martin–Löf's Type Theory*, (Clarendon Press, Oxford, 1990).
35. E. Palmgren: *An information system interpretation of Martin-Löf's partial type theory with universes*, Information and Computation1106 (1993) 26–60.
36. E. Palmgren: *On universes in type theory*. In: G. Sambin, J. smith (eds.): *Twenty-five Years of Type Theory* (Oxford University Press, Oxford, 1998) 191–204.
37. J. Paris, L. Harrington: *A mathematical incompleteness in Peano arithmetic*. In: J. Barwise (ed.): *Handbook of Mathematical Logic* (North Holland, Amsterdam, 1977) 1133–1142.
38. D. Prawitz: *Meaning and proofs: on the conflict between classical and intuitionistic logic.* Theoria 43 (1977)11–40.
39. M. Rathjen: *Proof-Theoretic Analysis of KPM*, Arch. Math. Logic 30 (1991) 377–403.
40. M. Rathjen: *The strength of some Martin–Löf type theories.* Preprint, Department of Mathematics, Ohio State University (1993) 39 pp.
41. M. Rathjen: *Proof theory of reflection.* Annals of Pure and Applied Logic 68 (1994) 181–224.
42. M. Rathjen: *The recursively Mahlo property in second order arithmetic.* Mathematical Logic Quarterly 42 (1996) 59–66.
43. M. Rathjen, E. Palmgren: *Inaccessibility in constructive set theory and type theory.* Annals of Pure and Applied Logic 94 (1998) 181–200.
44. M. Rathjen: *Explicit mathematics with the monotone fixed point principle. II: Models.* Journal of Symbolic Logic 64 (1999) 517–550.
45. M. Rathjen: *The superjump in Martin-Löf type theory.* In: S. Buss, P. Hajek, P. Pudlak (eds.): *Logic Colloquium '98*, Lecture Notes in Logic 13. (Association for Symbolic Logic, 2000) 363–386.
46. M. Rathjen: *Realizing Mahlo set theory in type theory*. Archive for Mathematical Logic 42 (2003) 89–101.
47. M. Rathjen: *The constructive Hilbert programme and the limits of Martin-Löf type theory* Synthese 147 (2005) 81–120.
48. B. Russell: *Mathematical logic as based on the theory of types*. American Journal of Mathematics 30 (1908) 222–262.
49. T. Setzer: *Proof theoretical strength of Martin–Löf type theory with W–type and one universe* (Thesis, University of Munich, 1993).
50. A. Setzer: *A well-ordering proof for the proof theoretical strength of Martin-Löf type theory*, Annals of Pure and Applied Logic 92 (1998) 113–159.
51. A. Setzer: *Extending Martin-Löf type theory by one Mahlo-universe*. Archive for Mathematical Logic (2000) 39: 155–181.
52. S. Simpson: *Nichtbeweisbarkeit von gewissen kombinatorischen Eigenschaften endlicher Bäume*, Archiv f. Math. Logik 25 (1985) 45–65.
53. S. Simpson: *Partial realizations of Hilbert's program.* Journal of Symbolic Logic 53 (1988) 349–363.
54. S. Simpson: *Subsystems of second order arithmetic* (Springer, Berlin, 1999).
55. W. Tait: *Finitism*. Journal of Philosophy 78 (1981) 524–546.
56. B. van der Waerden: *Moderne Algebra* I/II (Springer, Berlin, 1930/31)
57. H. Weyl: *Philosophy of Mathematics and Natural Sciences*. (Princeton University Press, Princeton, 1949)
58. W.H. Woodin: Large cardinal axioms and independence: The continuum problem revisited, The Mathematical Intelligencer vol. 16, No. 3 (1994) 31–35.

Categories, Structures, and the Frege-Hilbert Controversy: The Status of Meta-mathematics

Stewart Shapiro

The slogan of structuralism is that mathematics is the science of structure. Rather than focusing on the nature of individual mathematical objects, such as natural numbers, the structuralist contends that the subject matter of arithmetic, for example, is the structure of any collection of objects that has a designated, initial object and a successor relation that satisfies the induction principle. In the contemporary scene, Paul Benacerraf's classic "What numbers could not be" [2] provides the standard motivation for structuralism, arguing that numbers are not objects and that numerals are not singular terms. According to Benacerraf, elementary arithmetic is concerned with systems that share the common structure, and not with any particular ontology. As structuralism was gaining momentum in the philosophy of mathematics, Colin McLarty's "Numbers can be just what they have to" [19] put forward the thesis that category theory provides the proper, or at least an especially insightful and compelling, framework for it. He points out that objects (and arrows) in categories have only relational properties, which are just the features that the structuralist focuses on in systems that exemplify the natural number structure. Recent issues of this journal explore the relationship between category theory and structuralism. Geoffrey Hellman [14] challenges the foundational claims made on behalf of category theory, and McLarty [20] and Steve Awodey [1] reply. It is interesting that these two replies take the debate in competing and perhaps incompatible directions.

The title of Hellman [14] is a question: "Does category theory provide a framework for mathematical structuralism?", and Awodey's opening sentence provides an answer: "yes, obviously". Like just about everything else, it depends on what question is being asked. What sort of framework are we after? And what is mathematical structuralism? For that matter, what is category theory? I suspect that, to some extent at least, the various sides of this debate are at cross-purposes, focusing on different sets of issues and questions. Even if this is true, it does not follow that the debate is useless, uninteresting, or unimportant. It may be that the issues and questions of

S. Shapiro (✉)
The Ohio State University, OH, USA; University of St. Andrews, Scotland
e-mail: shapiro.4@osu.edu

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_18, © Oxford University Press 2005. Previously
published in *Philosophia Mathematica* 13:1, pp. 61–77. Published here with kind
permission.

one side are more central to mutual concerns than the issues and questions of the other side.

In this note, I hope to shed a little light on the question, or questions, by relating the present debate to a clash that took place over a hundred years ago, between two intellectual giants, Gottlob Frege and David Hilbert. I propose to focus on the role and function of meta-mathematics, which, I suggest, does not fit smoothly into Hilbert's algebraic perspective at the time. The problem was directly remedied in the subsequent development of the Hilbert program some decades later, where it is explicit that the proper meta-mathematics is finitary arithmetic. But, the story goes, this resolution was undermined with the incompleteness theorems, thanks to Gödel. So there is some unfinished business in the original debate, at least from Hilbert's side of it.

The general issue concerning meta-mathematics provides some perspective to the current debate over category theory and structuralism. The category theory folks, or at least some of them, are squarely on Hilbert's side of the Frege-Hilbert divide. This is no accident, of course. Saunders Mac Lane's roots go back to Göttingen. In part, Hellman's approach to structuralism [12, 13], as well as my own, also fall on Hilbert's side (see, for example, Shapiro ([24], Chapter 5, § 3) or Shapiro [23]). However, my own ante rem structuralism is an attempt to have it both ways. At least some of the questions answered in Shapiro [24] derive from what may be called the Frege-Quine tradition, and they concern the aforementioned unfinished business of the Hilbert program—the proper role or place of meta-mathematics. In my book, I use the word "structure" as a sortal, with quantifiers ranging over structures. I took this as a burden to say something about what a structure is, and I was led to traditional talk of universals and platonic forms. Fellow structuralists, such as Hellman [12, 13] and Michael Resnik [21], accept the same problematic, and give different answers than mine, the former being a structuralism without structures. The issue at hand is distinctly Fregean. We try to say what a structure is, and when two structures are identical or distinct. In short, we require a mathematical and/or a philosophical *theory* of structures or of systems-with-shared-structure. One of the burdens of McLarty [20] is to show that the meta-mathematical matters can themselves be approached from the categorical perspective.

The debate between Frege and Hilbert concerned geometry. Alberto Coffa ([6], 8, 17) provides a delightful summary of the situation on the ground at the time:

> During the second half of the nineteenth century, through a process still awaiting explanation, the community of geometers reached the conclusion that all geometries were here to stay . . . [T]his had all the appearance of being the first time that a community of scientists had agreed to accept in a not-merely-provisory way all the members of a set of mutually inconsistent theories about a certain domain . . . It was now up to philosophers . . . to make epistemological sense of the mathematicians' attitude toward geometry . . . The challenge was a difficult test for philosophers, a test which (sad to say) they all failed. . .
>
> For decades professional philosophers had remained largely unmoved by the new developments, watching them from afar or not at all . . . As the trend toward formalism became stronger and more definite, however, some philosophers concluded that the noble science of geometry was taking too harsh a beating from its practitioners. Perhaps it was time to take a stand on their behalf. In 1899, philosophy and geometry finally stood in eyeball-to-eyeball confrontation. The issue was to determine what, exactly, was going on in the new geometry.

What was going on, I believe, was that geometry was becoming less the science of space or space-time, and more the formal study of certain structures. Issues concerning the proper application of geometry to physics were being separated from the status of pure geometry, the branch of mathematics.[1] Hilbert's *Grundlagen der Geometrie* [15] represents the culmination of this development, delivering a death blow to a role for intuition or perception in the practice of geometry. Although intuition or observation may be the source of axioms, it plays no role in the actual pursuit of the subject.

The early pages of Hilbert [15] contain phrases like "the axioms of this group define the idea expressed by the word 'between' ..." and "the axioms of this group define the notion of congruence or motion". The idea is summed up as follows:

> We think of ... points, straight lines, and planes as having certain mutual relations, which we indicate by means of such words as "are situated", "between", "parallel", "congruent", "continuous", etc. The complete and exact description of these relations follows as a consequence of the *axioms of geometry*.

A crucial aspect of the axiomatization is that the system is what I call "free-standing". Anything at all can play the role of the undefined primitives of points, lines, planes, etc., so long as the axioms are satisfied. Hilbert was not out to capture the essence of a specific chunk of reality, be it space, the forms of intuition, or anything else. Otto Blumenthal reports that in a discussion in a Berlin train station in 1891, Hilbert said that in a proper axiomatization of geometry, "one must always be able to say, instead of 'points, straight lines, and planes', 'tables, chairs, and beer mugs'."[2]

In a retrospective encyclopedia article, Hilbert's student and protege Paul Bernays ([3], 497) summed up the aims of the new geometry:

> A main feature of Hilbert's axiomatization of geometry is that the axiomatic method is presented and practiced in the spirit of the abstract conception of mathematics that arose at the end of the nineteenth century and which has generally been adopted in modern mathematics. It consists in abstracting from the intuitive meaning of the terms ... and in understanding the assertions (theorems) of the axiomatized theory in a hypothetical sense, that is, as holding true for any interpretation ... for which the axioms are satisfied. Thus, an axiom system is regarded not as a system of statements about a subject matter but as a system of conditions for what might be called a relational structure ... [On] this conception of axiomatics, ... logical reasoning on the basis of the axioms is used not merely as a means of assisting intuition in the study of spatial figures; rather logical dependencies are considered for their own sake, and it is insisted that in reasoning we should rely only on those properties of a figure that either are explicitly assumed or follow logically from the assumptions and axioms.

---

[1] Coffa's claim that philosophers had ignored and then opposed the developments in geometry is a bit of an exaggeration. Husserl [18] made effective use of the new perspective in developing his metaphysics and philosophy of science (see Chapter 11, especially § § 70–71). Thanks to Per Martin-Löf for the reference. Coffa focuses on Frege and Russell.

[2] "Lebensgeschichte" in Hilbert ([17], 388–429); the story is related on p. 403.

At first, Frege had trouble with this orientation to mathematics. In a letter dated December 27, 1899, he lectured Hilbert on the nature of definitions and axioms.[3] According to Frege, axioms should express truths; definitions should give the meanings and fix the denotations of terms. These are fundamentally different enterprises, and should never be confused. Moreover, with a Hilbert-style implicit definition, *neither* job is accomplished. Frege complained that Hilbert [15] does not provide a definition of, say, "between" since the axiomatization "does not give a characteristic mark by which one could recognize whether the relation Between obtains":

> . . . the meanings of the words "point", "line", "between" are not given, but are assumed to be known in advance . . . [I]t is also left unclear what you call a point. One first thinks of points in the sense of Euclidean geometry, . . . But afterwards you think of a pair of numbers as a point . . . Here the axioms are made to carry a burden that belongs to definitions . . . [B]eside the old meaning of the word "axiom", . . . there emerges another meaning but one which I cannot grasp.

According to Frege, definitions are in sharp contrast with axioms and theorems. The latter

> . . . must not contain a word or sign whose sense and meaning, or whose contribution to the expression of a thought, was not already completely laid down, so that there is no doubt about the sense of the proposition and the thought it expresses. The only question can be whether this thought is true and what its truth rests on. Thus axioms and theorems can never try to lay down the meaning of a sign or word that occurs in them, but it must already be laid down.

Frege's point is a simple one. If the terms in the proposed "axioms" do not have meaning beforehand, then the statements cannot be true (or false), and thus they cannot be axioms. If they do have meaning beforehand, then the "axioms" cannot be definitions.

Hilbert replied on December 29, rejecting Frege's suggestion that the meanings of the words "point", "line", and "plane" are "not given, but are assumed to be known in advance":

> I do not want to assume anything as known in advance. I regard my explanation . . . as the definition of the concepts point, line, plane . . . If one is looking for other definitions of a "point", e.g. through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; one is looking for something one can never find because there is nothing there; and everything gets lost and becomes vague and tangled and degenerates into a game of hide and seek.

This talk of paraphrase is an allusion to "definitions" like Euclid's "a point is that which has no parts". Such "definitions" play no role in the mathematical development, and are thus irrelevant. Later in the same letter, when responding to the complaint that his notion of "point" is not "unequivocally fixed", Hilbert wrote:

> . . . it is surely obvious that every theory is only a scaffolding or schema of concepts together with their necessary relations to one another, and that the basic elements can be thought of

[3] The correspondence between Frege and Hilbert is published in Frege [9] and translated in Frege [10]. See Blanchette [5] for an insightful discussion of Frege's notion of logical consequence.

> in any way one likes. If in speaking of my points, I think of some system of things, e.g., the system love, law, chimney-sweep ... and then assume all my axioms as relations between these things, then my propositions, e.g., Pythagoras' theorem, are also valid for these things ... [A]ny theory can always be applied to infinitely many systems of basic elements. One only needs to apply a reversible one-one transformation and lay it down that the axioms shall be correspondingly the same for the transformed things. This circumstance is in fact frequently made use of, e.g. in the principle of duality ... [This] ... can never be a defect in a theory, and it is in any case unavoidable.

Note the similarity to the remark in the train station. Hilbert repeated the role of what is now called "implicit definition" (or, in philosophical circles, "functional definition") noting that it is impossible to give a definition of "point" in a few lines since "only the whole structure of axioms yields a complete definition". He noted the now familiar point that isomorphic structures are equivalent.

Frege did not get it, or did not want to. On the following September 16, he wrote that he could not reconcile the claim that axioms are definitions with Hilbert's view that axioms contain a precise and complete statement of the relations among the elementary concepts of a field of study. For Frege, "there can be talk about relations between concepts ... only after these concepts have been given sharp limits, but not while they are being defined". On September 22, an exasperated Hilbert replied:

> ... a concept can be fixed logically only by its relations to other concepts. These relations, formulated in certain statements I call axioms, thus arriving at the view that axioms ... are the definitions of the concepts. I did not think up this view because I had nothing better to do, but I found myself forced into it by the requirements of strictness in logical inference and in the logical construction of a theory. I have become convinced that the more subtle parts of mathematics ... can be treated with certainty only in this way; otherwise one is only going around in a circle.

Hilbert's claim that a concept can be fixed only by its relations to other concepts is a standard motivation for structuralism.[4]

Nowadays we have a rough and ready distinction which we can apply here. The algebraist says that a group is anything that satisfies the axioms of group theory; a ring is anything that satisfies the ring axioms, etc. There is no such thing as "*the* group" or "*the* ring". Hilbert says the same thing about geometry, and, by extension, arithmetic, real analysis, and so forth. At the time, it seems, Hilbert took *every* branch of mathematics to be algebraic: any given branch is "about" any system that satisfies its axioms. In opposition to this, Frege insisted that arithmetic and geometry each have a *specific* subject matter, space in the one case and the realm of natural numbers in the other. And the axioms express (presumably self-evident) truths about this subject matter. Following a suggestion of Hellman's, let us say that for Frege, the axioms of arithmetic and geometry are *assertory*; and for Hilbert,

---

[4] The exegetical and historical issues are complex, and it would take us too far afield to go much deeper. Did Hilbert intend his remark to be limited to mathematical, or perhaps just geometrical concepts? Consider, moreover, Frege's ([7], Introduction) own context principle that one can "never ask for the meaning of a word in isolation, but only in the context of a proposition". This can be, and has been, interpreted to entail that a concept can only be fixed by its relations to other concepts (see, for example, the neo-logicists Wright [27], Hale [11]).

they are *algebraic*. Sentences that are assertory are meant to express propositions with fixed truth values. Algebraic sentences are schematic, applying to any system of objects that meets certain given conditions.[5]

For what it is worth, my ante rem structuralism proposes to bridge the gap between the algebraic and assertory approach to theories like Euclidean geometry, arithmetic, real analysis and complex analysis. From one perspective, called "places-are-offices", the theories are algebraic, applying to whatever systems satisfy them. However, if the axioms of a branch of mathematics are satisfiable and categorical, then they characterize a (single) structure, and the axioms are true *of it*. I call this the "places-are-objects" perspective. The idea is that places in a structure are bone fide objects, and we can have quantifiers ranging over them. The structure itself is a chunk of reality, and the theory is about it. So the same axioms are algebraic from one perspective, and assertory from another ([24], Chapter 3).[6]

One can take any algebraic sentence and interpret it directly as a proposition about all systems of a certain sort. Consider, for example, the Euclidean sentence that there is a point that lies between any two distinct points. From the algebraic perspective, this comes to something like this:

> (*) In any (possible) Euclidean system *S*, for any two distinct objects *a, b* in *S* that are "points-in-*S*", there is a third object *c* that is also a point-in-*S*, and *c* lies between-in-*S a* and *b*.

This is the route of eliminative structuralism ala Benacerraf and modal structuralism ala Hellman, the latter supplying the "translation" schemes explicitly, with full mathematical rigor. The above passage from Bernays contains a sentence in this form, and such statements are at least implicit in Hilbert's motivating remarks and in the correspondence with Frege.

What is the status of statements like (*)? It would seem that for the algebraist, such sentences must themselves be assertory. This is just to insist that a philosopher or mathematician assert something when stating the algebraic position. Moreover, it would run counter to the spirit of Hilbert's approach to think of the opening quantifier in (*) as itself restricted to a particular system. A system of what? A system of systems? At least prima facie, then, if an algebraist insists that all (legitimate) mathe-

---

5 The word "algebraic" might be a bit misleading. There is a three-fold distinction that can be made here. Say that a theory is "Fregean" if it is intended to be about a specific subject matter, and that a theory is "Hilbertian" if it consists of taking the logical consequences of an axiomatization regarded as an implicit definition of a type of structure. Contemporary group theory and ring theory are not pursued, for more than a few minutes, in this Hilbertian manner. Rather, the group theorist studies all groups, developing relationships between them and with other structures. This study is made in a background framework, perhaps naive set theory, and one can take either a Fregean or a Hilbertian approach to this background.

6 It was thus potentially misleading for me to refer to theories like arithmetic and real analysis as "non-algebraic" in the motivating sections of Shapiro [24] (e.g., pp. 40–41). I took the distinction to be implicit in mathematical practice, although this begs the question against a thoroughgoing algebraist like the Hilbert of 1900. The correct idea, I think, is that theories like arithmetic and real analysis can be treated as assertory from one perspective, that of places-are-objects. This does not preclude them from also being treated algebraically.

matical statements are algebraic, then (*)-type assertions are not mathematical. But this seems ad hoc. In typical cases, the (*)-type assertions contain no non-logical terminology, and so for the algebraist, there is nothing to reinterpret. In an attempt to recapture Hilbert's perspective, Frege [8] himself showed how to make statements like (*) in his own logical system. And, of course, for Frege such statements, like all others in mathematics, are assertory.

Hilbert's thoroughgoing algebraic perspective is reminiscent of the oft-heard claim that a category is anything that satisfies the axioms of category theory, what Awodey calls the top-down approach. Contrary to what I once wrote ([24], 193), the "arrows" of a category do not have to be functions (as those notions are understood in set theory—not to quibble over terminology). The category theorist claims that her account, in terms of the axioms of category theory, is a more fruitful way to define and organize (algebraic) mathematics than Hilbert's method of implicit definition in higher-order languages (which is closer to the techniques of Shapiro [24]). I do not have the expertise to shed light on that matter, and it does not strike me as particularly philosophical. I see how many of the philosophical claims in my book can be formulated in terms of category theory, rather than the quasi-model-theoretic perspective I took at the time. I will not speak for my fellow structuralists.

Some of the central points in Hellman [14] have roots in a closely related matter on which Hilbert and Frege never saw eye to eye: the role of consistency and the nature of mathematical existence. Hilbert's *Grundlagen* provided consistency and independence proofs by finding interpretations that satisfy various sets of axioms. Typically, he would interpret the axioms of a theory in terms of constructions on real numbers. This approach, now as common as anything in mathematics, runs roughshod over Euclid's definition of a "point" as "that which has no parts". When we interpret a "point" as an ordered pair of real numbers, we see that points can indeed have parts. This free reinterpretation of axioms is a main strength of contemporary mathematical logic, and a mainstay of mathematics generally. It drives the structuralist, algebraic perspective on mathematics. And it runs counter to the Fregean perspective. In the first letter Frege complained that Hilbert takes "a pair of numbers as a point" contrasting this with "points in the sense of Euclidean geometry".

What is the Hilbertian to make of the statements of consistency themselves? Are they algebraic or assertory, or both at once? In the 1899 letter, Frege said that there is never a serious question of consistency. From the truth of axioms, concerning their intended subject matter, "it follows that they do not contradict one another". Since Hilbert did "not want to assume anything as known in advance", he rejected Frege's claim that we can reason from truth to consistency. He wrote:

> As long as I have been thinking, writing and lecturing on these things, I have been saying the exact reverse: if the arbitrarily given axioms do not contradict each other with all their consequences, then they are true and the things defined by them exist. This is for me the criterion of truth and existence.

On the algebraic, top-down approach, we characterize a structure and thereby study systems that exemplify that structure. Clearly, if the characterization is not coherent, then one has not characterized a possible system, and the enterprise has misfired. The more controversial claim is the converse: if the given characterization is coherent, then all has gone well. There is no further mathematical or metaphysical burden to discharge. The question of coherence is all that remains of the traditional issue of existence, at least for mathematics.

But what *of* this notion of coherence? Is it a *mathematical* question, and, if so, how do we negotiate it? For Hilbert, coherence is consistency, and by this he surely meant deductive, proof-theoretic consistency. If it is not possible to derive a contradiction from a collection of axioms, then the "the things defined by" the axioms exist, and the axioms "are true" of those things.

It seems clear that for Hilbert and just about anyone else, consistency is itself a mathematical matter. His methodology indicates that in order for us to be assured that certain mathematical objects exist, we have to establish the consistency of an axiomatization. In the *Grundlagen*, Hilbert discharged this burden, at least in part, by providing relative consistency *proofs*. For example, in showing how to interpret the axioms of a non-Euclidean geometry in the real numbers, he established that the non-Euclidean theory is consistent if real analysis is.

We now enter the realm of meta-mathematics. Given the way this matter is handled in Hilbert's *Grundlagen*, it is clear that meta-mathematics is itself mathematics. What are we to make of it? What is the status of *statements* (and proofs) of consistency? Are they assertory or algebraic?

This matter is not treated explicitly in Hilbert's *Grundlagen*, and it is hard to be definitive on what his view was, or should have been, but I suggest that the meta-theory—the mathematical theory in which the consistency of an axiomatization is established—is not to be understood algebraically, not as another theory of whatever satisfies *its* axioms. Instead, the statement that a given theory, such as Euclidean geometry, is consistent is itself assertory. The notion of consistency is a contentful property of theories, and is not to be understood as defined implicitly by the axioms of the meta-theory. For one thing, the meta-theory is not axiomatized in the *Grundlagen*, and so there is no implicit definition of the meta-theoretic notions. This, of course, is not decisive. It would be a routine exercise for a graduate student in mathematical logic to axiomatize the meta-theory of the *Grundlagen*. Given the structural analogy between natural numbers and strings, the meta-theory would resemble elementary arithmetic. However, if a Hilbertian algebraist did think of the axiomatized meta-theory as algebraic, then she would have to worry about *its* consistency. How would we establish that? The ensuing regress is vicious to the epistemological goals of the *Grundlagen*.[7]

In the later Hilbert program (e.g., [16]) relative consistency gives way to absolute consistency. There, the meta-theory is finitary proof theory, focused directly on formal languages themselves. It is explicit that finitary proof theory is not just the study of another structure, on a par with geometry and real analysis. Finitary proof theory has its own, unique subject matter, related to natural numbers and formal syntax,

[7] I am indebted to an anonymous referee for this point.

and it is ultimately founded on something in the neighborhood of Kantian intuition. Hilbert said that finitary proof theory is contentful. In present terms, the theorems of finitary proof theory are assertory, not algebraic.

Of course, thanks to Gödel, finitary proof theory proved to be all but useless for establishing consistency. We have come to live without absolute consistency proofs. The crisis in set theory passed, and we now fly without a Hilbert-style safety net. But, of course, the powerful algebraic orientation to mathematics continues to thrive today, as well it should. The orientation is championed by, among others, the advocates of category theory. But, in practice, how do we satisfy ourselves that a given characterization—whether it is a traditional axiomatization or a type of category—is coherent, and thus characterizes a structure or a possible system? This is the unfinished business from the Hilbert program that I mentioned earlier.

Here is the question: Do we still need some sort of meta-mathematics to answer legitimate foundational questions? If we do need meta-mathematics, can it be understood algebraically, or structurally, on a par with (pure) geometry, and, for the Hilbert-style algebraist, infinitary arithmetic and real analysis? Or is meta-mathematics an exception to the slogan that mathematics concerns structure?

As noted above, in the debate with Frege, Hilbert said that (deductive) consistency is sufficient for "existence", or, better, that consistency is all that remains of the traditional, metaphysical matter of existence. This much continued into the Hilbert program. If we restrict ourselves to first-order axiomatizations, then Gödel's completeness theorem does assure is that consistency implies existence. The theorem is that if a first-order axiomatization is consistent, then it has a model: there is a system that makes the axioms true. So perhaps Hilbert's claim about consistency foreshadowed the completeness theorem. But then what is the status of the completeness theorem itself? If the algebraist wants to use it to bolster the claim that consistency is all that remains of the traditional question of existence, then she must think of the completeness theorem as itself assertory. Indeed, she might *assert* it in defense of her algebraic claim. To play the foundational role, the meta-theory in which the completeness theorem is proved cannot be algebraic. It must be contentful.

Recall also that Hilbert's original axiomatizations are not first-order. Indeed, when first-order logic was originally separated out for special study, it was called the *restricted* functional calculus. The completeness theorem fails in higher-order systems (see [22], Chapter 4): deductive consistency does not entail coherence. Consider, for example, the theory I'll call PA-weird, consisting of the second-order Peano axioms together with the formal statement that second-order PA is *in*consistent. PA-weird is consistent, but it has no models and, arguably, it does not describe a coherent structure. In my structuralism book, I propose that coherence is to be modeled by satisfiability. But satisfiability seems to require an assertory theory of sets or structures.

This is not just an obscure matter for philosophers of mathematics. The practice of mathematics occasionally runs into serious questions concerning coherence, and, thus for the Hilbert-style algebraist or the structuralist, the existence of certain structures. Mark Wilson ([26], § III) illustrates the historical development and acceptance of a space-time with an "affine" structure on the temporal slices:

> ... the acceptance of ... non-traditional structures poses a delicate problem for philosophy of mathematics, viz., how can the novel structures be brought under the umbrella of *safe* mathematics? Certainly, we rightly feel, after sufficient doodles have been deposited on coffee shop napkins, that we understand the intended structure ... But it is hard to find a fully satisfactory way that permits a smooth integration of non-standard structures into mathematics ... We would hope that "any coherent structure we can dream up is worthy of mathematical study ... " The rub comes when we try to determine whether a proposed structure is "coherent" or not. Raw "intuition" cannot always be trusted; even the great Riemann accepted structures as coherent that later turned out to be impossible. *Existence principles* beyond "it seems okay to me" are needed to decide whether a proposed novel structure is genuinely coherent ... [L]ate nineteenth century mathematicians recognized that ... existence principles ... need to piggyback eventually upon some accepted range of more traditional mathematical structure, such as the ontological frames of arithmetic or Euclidean geometry. In ... our century, set theory has become the canonical backdrop to which questions of structural existence are referred. (pp. 208–209)

Within the community of professional mathematicians, if not philosophers, a set-theoretic proof of satisfiability resolves any legitimate questions of existence.

But we now have an especially sharp version of the previous question concerning meta-theory. What are we to make of the set-theoretic model theory used to resolve questions of coherence in practice? Set theory, of course, is far more substantial than finitary proof theory, but it plays a similar role in adjudicating matters of "existence". There are two theoretical options concerning the meta-mathematical background, what Wilson calls the "canonical backdrop". One is to argue that for set theory to play the foundational role, it is *not* to be understood algebraically. On this view, set theory has a subject matter, the iterative hierarchy V. It is an *assertory* theory about how various structures relate to, and interact with, each other. Our first option, then, is to think of the background meta-mathematics—model theory—in the same way that Frege thought of arithmetic and geometry, and the same way that Hilbert understood finitary proof theory and, arguably, the assertions of relative consistency in his *Grundlagen der Geometrie*.

This orientation toward the meta-theory supports Hellman's [14] claim that the category theorist requires an "external" theory of relations, functions, and the like. This "home address" issue concerns the nature of the fundamental terminology of category theory. Hellman's point is that the talk of relations and functions (or functors) in the informal language of category theory must be assertory, and that set theory is a natural background for such assertions. It is not that one thinks of the iterative hierarchy as literally containing all structures, or all categories. Rather, we think of the iterative hierarchy as containing an isomorphism type for each structure. My own structure theory ([24], Chapter 3, § 4) was meant to play the same assertory, foundational role as set theory, and, indeed, structure theory is a notational variant of set theory. Hellman's [12] modal set theory does the same work, without presupposing the (actual) existence of any abstract objects. In present terms, the point here is that the meta-theory, whatever it is, must itself be assertory, and thus an exception to the slogan that mathematics is the science of structure.

A category theorist who goes for this first theoretical option concerning the meta-theory is not without resources. McLarty [20] argues that a set theory formulated in categorical terms, such as ETCS or CCAF, will work even better as a canonical backdrop than the more standard Zermelo-Fraenkel set theory. To be sure, if a category-based theory is to play this role, then *its* axioms must be assertory. The canonical backdrop, whatever it may be, is "external" to the algebraic perspective. But there is nothing to prevent the category-theorist from understanding a theory as assertory. This is not to say that McLarty himself takes the axioms of some category-based set theory this way. His point is that if one needs or wants a set theory to serve as canonical backdrop for questions of existence, as on our first option, then a category-based set theory will do the job as good as, or better than, the more standard Zermelo-Fraenkel set theory. I must plead ignorance, or at least incompetence, concerning the proper mathematical theoretical framework for such meta-mathematical questions, once it is agreed that such questions are legitimate, and assertory. The various category theories and set theories are inter-translatable, and the debate sometimes turns on which is more natural.

In any case, our first theoretical option is not quite the bottom-up approach alluded to in Awodey [1]. The algebraic structuralist does not *construct* the structures of mathematics within his or her favored set theory. Set theory does not supply the ultimate subject matter for any branch of mathematics. Rather, we use set theory to establish that a given theory is coherent.

In the account of ante rem structuralism in Shapiro [24], I said that a structure *consists* of places and relations. So far, this is only a structural claim, analogous to the set-theoretical thesis that a set consists of its members. However, I provided a mathematical theory of structures and their places, and I suppose that I was thinking of the subject matter of that theory as the universe of all of mathematics. In that sense, my account is bottom-up, within what we may call the Frege-Quine tradition. I presume that if ordinary model theory is understood as a semantics for mathematical languages, it too is bottom-up in the same sense, since it aims to provide a theoretical account of the interpretations of various theories. In any case, there is little need to quibble over labels.

To sum up our first theoretical option, standard set theory, the category-based set theory suggested by McLarty, my own structure theory, or Hellman's modal set theory, are themselves assertory theories of sets, structures, the possible existence of systems, etc. As such, each of them is not just another mathematical theory, providing an implicit definition of some structures, or isomorphism types. The reason for this is that set theory, structure theory, etc., has a foundational role to play concerning the coherence of definitions. And this last is an assertory matter.

Anyway, this is one way to look at the meta-mathematical background. A second theoretical option, more in line with Awodey [1], is to kick away the foundational ladder altogether, and take the meta-mathematical set theory, structure theory, or whatever, itself to be an algebraic theory. On this view, set theory does not directly serve as a court of appeal for matters of coherence and thus existence, neither in the sense of supplying the ultimate ontology for mathematics, on the bottom-up approach, nor in the attenuated sense of supplying isomorphism types of everything in mathematics. The axioms of set theory are just implicit definitions that, if coherent, characterize a structure or a class of structures. The same goes for structure theory, modal set theory, and the various topos theories.

On this view, everything in mathematics is algebraic. So if there is to be an assertory canonical backdrop—a non-algebraic theory of coherence, consistency, mathematical existence, whatever—it will be relegated outside of mathematics, perhaps to philosophy. But this seems an ad hoc way to draw boundaries between disciplines. As we have seen, at least some questions of existence have been settled, by mathematicians, via rigorous proof. So this is a tough pill to swallow.

On our second theoretical option, set theory may still help when serious questions of coherence arise in practice for theories other than set theory. We go back to the plan executed in Hilbert's *Grundlagen*, and settle for the analogue of relative consistency proofs. By finding a model of a given theory T in set theory, or in "the" iterative hierarchy, one shows that T is coherent, and defines a structure, *if set theory does*. Such a proof is effective whether the background set theory is formulated in traditional terms or in the idiom of category theory, as suggested in McLarty [20].

Notice that we have no formal assurance that our set theory is itself coherent, and thus characterizes a structure or possible system. But perhaps we don't need such assurance. On the first theoretical option, where the meta-theory is assertory, we likewise have no theoretical assurance that set theory is *true*. Again, we have no safety net, and do not really need one.

From the present, thoroughly algebraic perspective, we are still left with the notion of relative consistency. And our nagging question returns. What are we to make of a statement of relative consistency—a proposition that a given theory T is consistent if set theory is? Since, on the view in question, all of mathematics is algebraic, there is no room for any assertory matters in mathematics. So if we insist on the letter of our second theoretical option, and if statements of relative consistency are assertory, as seems obvious, then they are not mathematical. But it seems equally obvious that relative consistency is a mathematical matter.

Perhaps at least one of these "obvious" theses can be resisted. On our second theoretical option, to speak *mathematically* about consistency, coherence, or the like, is just to speak within an algebraic theory. The statements of relative consistency hold in any interpretation or reinterpretation of the non-logical terminology in the background set theory, which includes things like the membership symbol and a sign for satisfaction. Just as (pure) geometry is not *about* any particular things, be they space points or space-time points, (pure) set theory, proof theory, model theory, formal semantics, and the like, are not about any particular things, be they theories, consistency, deductions, interpretations, models, or the like. To be sure, we do make assertory statements about physical space—in physics itself for example. Such statements are part of the standard *application* of geometry. In doing physical geometry, we interpret the non-logical terminology of pure geometry accordingly, and the axioms, so interpreted, may or may not be true. Similarly, assertory statements about interpretations, deductions, relative consistency, and the like, are an application of the background meta-theory, perhaps the standard application. Just as mathematics became liberated from intuitions concerning physical space and space-time in the nineteenth century, contemporary mathematical logic is similarly liberated from theories, interpretations, deductions, and consistency. As a structuralist, our theorist can hold—in assertory philosophy—that satisfiability, consistency, or coherence implies existence, but she cannot maintain that any of these notions are mathematical matters. There simply are no distinctly mathematical objects, and so theories, deductions, and interpretations are not mathematical. But perhaps we should not quibble over labels.[8]

This is not the place to decide which of the theoretical options are best for structuralism, or for category theory, or for mathematics generally. In this note, I will rest content if I have managed to sharpen the battle lines a little.

Acknowledgement This article was first published in *Philosophia Mathematica (3) 13* (2005), 61–77. The editors, the publisher, and I are most grateful to Oxford University Press for generous permission to reprint the article here. I am indebted to Steve Awodey, Geoffrey Hellman, and Robert Thomas for helpful discussion on a previous version of this article. I am especially indebted to the referees for *Philosophia Mathematica*, one of whom provided a long collection of insightful comments on an earlier draft. Thanks also to the audience at a session on category theory and structuralism, held at the 2004 Spring meeting of the Association for Symbolic Logic in Chicago, and to the audience at a workshop on the foundations of mathematics, held in Uppsala, Sweden, in August 2004, which gave rise to this volume.

References

1. Awodey, S. [2004], “An answer to Hellman’s question: Does category theory provide a framework for mathematical structuralism?”, *Philosophia Mathematica (3) 12*, 54–64.
2. Benacerraf, P. [1965], “What numbers could not be”, *Philosophical Review 74*, 47–73; reprinted in [4].
3. Bernays, P. [1967], “Hilbert, David” in *The encyclopedia of philosophy, Volume 3*, edited by P. Edwards, New York, Macmillan publishing company and The Free Press, 496–504.
4. Benacerraf, P., and H. Putnam [1983], *Philosophy of mathematics*, second edition, Cambridge, Cambridge University Press.
5. Blanchette, P. [1996], “Frege and Hilbert on consistency”, *Journal of Philosophy 93*, 317–336.
6. Coffa, A. [1986], “From Geometry to tolerance: sources of conventionalism in nineteenth-century geometry” in *From quarks to quasars: Philosophical problems of modern physics*, University of Pittsburgh Series, Volume 7, Pittsburgh, Pittsburgh University Press, 3–70.

---

[8] On one occasion when I presented an ancestor of this paper, a mathematician in the audience endorsed the thoroughgoing algebraic perspective, calling it a “Copernican revolution”. Just as the pioneers of modern physics and astronomy showed that we do not have to think of the earth as unmovable, or even unmoving, so the pioneers of modern mathematics showed that we do not have to see talk of numbers and the like as grounded in a solid subject matter. It is an intriguing analogy, but I am not sure what was meant. If the idea is that the modern mathematician does not need absolute certainty that her axioms are consistent, then it applies just as well to the less than thoroughgoing perspective of our first option. It is just the lack of a safety net, noted above. However, if the mathematician’s comment was meant to say that (pure) mathematics and metamathematics can get by without making any assertions at all, then it is indeed our second option. For what it is worth, a number of other mathematicians and philosophers in the audience expressed bafflement at the second option.

7. Frege, G. [1884], *Die Grundlagen der Arithmetik*, Breslau, Koebner; *The foundations of arithmetic*, translated by J. Austin, second edition, New York, Harper, 1960.
8. Frege, G. [1971], *On the foundations of geometry and formal theories of arithmetic*, translated by Eikee-Henner W. Kluge, New Haven, Connecticut, Yale University Press.
9. Frege, G. [1976], *Wissenschaftlicher Briefwechsel*, edited by G. Gabriel, H. Hermes, F. Kambartel, and C. Thiel, Hamburg, Felix Meiner.
10. Frege, G. [1980], *Philosophical and mathematical correspondence*, Oxford, Basil Blackwell.
11. Hale, B. [1987], *Abstract objects*, Oxford, Basil Blackwell.
12. Hellman, G. [1989], *Mathematics without numbers*, Oxford, Oxford University Press.
13. Hellman, G. [1996], "Structuralism without structures", *Philosophia Mathematica (III) 4*, 100–123.
14. Hellman, G. [2003], "Does category theory provide a framework for mathematical structuralism?", *Philosophia Mathematica (3) 11*, 129–157.
15. Hilbert, D. [1899], *Grundlagen der Geometrie*, Leipzig, Teubner; *Foundations of geometry*, translated by E. Townsend, La Salle, Illinois, Open Court, 1959.
16. Hilbert, D. [1925], "Über das Unendliche", *Mathematische Annalen 95*, 161–190; translated as "On the infinite", [25], 369–392 and [4], 183–201.
17. Hilbert, D. [1935], *Gesammelte Abhandlungen, Dritter Band*, Berlin, Julius Springer.
18. Husserl, E. [1900], *Logische Untersuchen 1*, Halle a.d. S., M. Niemeyer, translated as *Logical Investigations*, by J. N. Findlay, Amherst, New York, Humanity Books, 2000.
19. McLarty, C. [1993], "Numbers can be just what they have to", *Nous 27*, 487–498.
20. McLarty, C. [2004], "Exploring categorical structuralism", *Philosophia Mathematica (3) 12*, 37–53.
21. Resnik, M. [1997], *Mathematics as a science of patterns*, Oxford, Oxford University Press.
22. Shapiro, S. [1991], *Foundations without foundationalism: A case for second-order logic*, Oxford, Oxford University Press.
23. Shapiro, S. [1996], "Space, number, and structure: A tale of two debates", *Philosophia Mathematica (3) 4* (1996), 148–173.
24. Shapiro, S. [1997], *Philosophy of mathematics: structure and ontology*, New York, Oxford University Press.
25. Van Heijenoort, J. [1967], *From Frege to Gödel*, Cambridge, Massachusetts, Harvard University Press.
26. Wilson, M. [1993], "There's a hole and a bucket, dear Leibniz", *Midwest Studies in Philosophy 18*, 202–241.
27. Wright, C. [1983], *Frege's conception of numbers as objects*, Aberdeen, Aberdeen University Press.

Beyond Hilbert's Reach?

Wilfried Sieg

...historical reflection serves, in the end, to shape the present and the future.

Ernst Cassirer[1]

Abstract Work in the foundations of mathematics should provide systematic frameworks for important parts of the practice of mathematics, and the frameworks should be grounded in conceptual analyses that reflect central aspects of mathematical experience. The Hilbert School of the 1920s used suitable frameworks to formalize (parts of) mathematics and provided conceptual analyses. However, its analyses were mostly restricted to finitist mathematics, the programmatic basis for proving the consistency of frameworks and, thus, their instrumental usefulness. Is the broader foundational quest beyond Hilbert's reach? The answer to this question seems simple: "Yes & No". It is "Yes", if we focus exclusively on Hilbert's finitism; it is "No", if we take into account the more sweeping scope of Hilbert and Bernays's foundational thinking. The evident limitations of Hilbert's "formalism" have been pointed out all too frequently; in contrast, I will trace connections of Hilbert's work, beginning in the late 19th century, to contemporary work in mathematical logic. Bernays's reflective philosophical investigations play a significant role in reinforcing these connections. My paper pursues two complementary goals, namely, to describe a global, integrating perspective for foundational work and to formulate some more local, focused problems for mathematical work.

1 What is at Issue?

It is a fact of intellectual history, perhaps a curious one, but nonetheless a fact, that the *Grundlagenstreit* of the 1920s colors even now our perspectives on the foundations of mathematics and beyond. In those early debates, we find dramatically formulated stances, and we tend to interpret them as being substantively and starkly

---

To Howard Stein – friend and teacher.

[1] See [58], p. 112. "Immer wieder schärft er seinen Zuhörern ein, da alle historische Betrachtung letztlich - symbolisch - im Dienste der Gestaltung der Gegenwart und der Zukunft steht."

W. Sieg (✉)
Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: Sieg@CMU.EDU

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341,
DOI 10.1007/978-1-4020-8926-8_19,

opposed to each other. Minimal historical awareness should have undermined that tendency a long time ago, as the finitist program, first formulated in Hilbert's Leipzig talk of September 22, 1922 [45], was explicitly intended to mediate between constructivist and logicist, set theoretic positions. Weyl recognized that point in papers[2] from 1925 [87] and 1928 [88], and even Brouwer's polemical *Intuitionistic reflections on formalism*, published in 1927 [16], contain these remarks:

> The disagreement over which is correct, the formalistic way of founding mathematics anew or the intuitionistic way of reconstructing it, will vanish, and the choice between the two activities be reduced to a matter of taste, as soon as the following insights[3] ... are generally accepted. The acceptance of these insights is only a question of time, since they are the results of pure reflection and hence contain no disputable element, so that anyone who has once understood them must accept them.[4]

From Hilbert's perspective, there was every principled reason to view the mathematical substance of Weyl's and Brouwer's foundational work as part of broadly conceived axiomatic investigations; in lecture notes from the summer term of 1920 (*Probleme der mathematischen Logik*; unpublished lecture note) he had already emphasized:

> But what these two researchers [Weyl and Brouwer] have achieved in terms of positive and fruitful results through their investigations on the foundations of mathematics, that fits very well with the axiomatic method, indeed, it is exactly in the spirit of this method. For it is being investigated, how a part of analysis can be delimited by a narrower system of assumptions.[5]

One is tempted to think that the *Grundlagenstreit* could have given way to a calmer discussion; but it did not.[6] I will not try to disentangle aspects of

---

[2] Cf. p. 540 ff. in Weyl [87], respectively pp. 482–4 in Weyl [88].

[3] Brouwer lists four basic insights. The first two are constitutive of Hilbert's proof theoretic program, and Brouwer emphasizes that they have been "understood and accepted in the formalistic literature", not without claiming that they have been taken over – without proper acknowledgement – from intuitionism. The first insight concerns the distinction between the construction of formalistic mathematics and the intuitive (contentual) metamathematics concerning this construction; the second insight points to the problematic character of the law of excluded middle (lem). The remaining two insights are formulated straightforwardly and with great clarity, namely, that the lem is identical with the claim that every mathematical problem is solvable, and that consistency does not guarantee correctness. The fourth insight can be reformulated as stating that consistency does not provide a contentual justification of formalistic mathematics. The third point would have been disputed, and the fourth was in this general formulation undoubtedly clear to Hilbert through his early work on Non-Euclidean geometries. Brouwer made also a very specific claim concerning the lem as part of the fourth insight, namely, that its correctness can be justified only by the lem itself. This claim was taken back in Brouwer [18], p. 14, fn. 1; cf. the editor's addition to footnote 8 in Brouwer [16] on p. 460 of (van Heijenoort). The substance of the claim had been refuted already earlier by the Gödel-Gentzen reduction of classical to intuitionistic arithmetic.

[4] See [17], p. 490.

[5] Hilbert 1920 (unpublished), p. 34. "Was aber diese beiden Forscher [Weyl and Brouwer] in ihren Untersuchungen über die Grundlagen der Mathematik an Positivem und Fruchtbarem leisten, das fügt sich der axiomatischen Methode durchaus ein und ist gerade im Sinne dieser Methode. Denn es wird hier untersucht wie sich ein Teil der Analysis durch ein gewisses engeres System von Voraussetzungen abgrenzen lässt."

[6] It should be noted, however, that some literally identified finitism and intuitionism - before Gödel's and Gentzen's result. I am thinking of Bernays, von Neumann, and Herbrand "within"

personality, professional aspirations, or ideological judgements. Rather, I intend to describe the broad foundational context and attempt to understand, what it is the programmatic constructivist Hilbert defended against his flamboyant fellow constructivist Brouwer, how he tried to do so, and how he got there.

The logico-philosophical community has focused on Hilbert's finitist means for securing "classical" mathematics and on the epistemological distinctiveness of those means - as viewed in the twenties, and I think that a deepened mathematical and philosophical analysis of finitism remains an important issue. However, we have not been equally concerned with the substance of what Hilbert strove to secure - over a lifetime. And that is not *classical* mathematics as it evolved until the 19th century, but rather *modern* mathematics as it resulted from a radical transformation during the second half of that century. Howard Stein called it a transformation "so profound that it is not too much to call it a second birth of the subject"; he argued, and I agree, that it was effected mainly by the work of Gauss, Dirichlet, Riemann, and Dedekind.[7]

Hilbert was intimately connected to this part of mathematical tradition (in Göttingen), but also to a second significant aspect. I am alluding to the free use of mathematical concepts in, and indeed their invention or free creation for, applications in the sciences. (I should point out that Weyl, in 1928 [88], viewed intuitionistic mathematics as inadequate for the sciences!) In the introductory remarks of his Paris Lecture Hilbert described most vividly the rich interplay of mathematical thought and experience. Discussing the central importance of problems for mathematics, he commented on their origins as follows:

> ... Surely the first and oldest problems in every branch of mathematics spring from experience and are suggested by the world of external phenomena. ... But, in the further development of a branch of mathematics, the human mind, encouraged by the success of its solutions, becomes conscious of its independence. It evolves from itself alone ... by means of logical combination, generalization, specialization, ... and appears then itself as the real questioner.. . .
>
> ... while the creative power of pure reason is at work, the outer world again comes into play, forces upon us new questions from actual experience, opens up new branches of mathematics .... And it seems to me that the numerous and surprising analogies and the apparent harmony which the mathematician so often perceives in the questions, methods, and ideas of the various branches of his science, have their origin in this ever-recurring interplay between thought and experience (Whitehead and Russell [89]).[8]

One basic condition has to be met, however, in order to safeguard creative freedom within mathematics and within contexts of applications: the introduced notions must be consistent. That was clearly expressed in the Paris Lecture, and Hilbert reiterated this view four years later in his Heidelberg talk. About the underlying

---

the Hilbert school, but also of Carnap and Fraenkel. As to the former, I am alluding to Carnap [21], in particular, pp. 309–10; as to the latter, let me quote the ironic (but historically inaccurate) remark in his (1930) [32], p. 294, "Wie mir scheint, hat *Brouwer* den gröten Erfolg für seine Anschauungen dadurch erzielt, da er als Anhänger seiner Ausgangsposition – *Hilbert* gewonnen hat!"

7 See [82], p. 238.

8 See [44], p. 1098.

creative principle he wrote then: "... in its freest use [that principle] justifies us in forming ever new notions with the sole restriction that we avoid a contradiction".[9] Thus, the central methodological issue is, how we can rationally assess whether this restriction has been met.

The methodological issue was more concrete and limited already at this point: Hilbert sought to establish the consistency of axiom systems, e.g., in (1900) [42] for the real numbers and in (1904, *Zahlbegriff und Quadratur des Kreises*; unpublished lecture note) for the natural numbers. Such a proof was to ensure the existence of the set or, in Cantor's terminology, of the consistent multiplicity of the real and natural numbers. The issue can be traced back to Dedekind and is, according to Bernays (and in harmony with my earlier remarks), most closely connected to the "transformation the methodological approach of mathematics underwent towards the end of the 19th century".[10] One characteristic feature of that transformation is the emergence of *existential axiomatics* described in the first part of my paper. That part is entitled "Logical Models" and examines Dedekind's and Hilbert's attempts to secure the consistency of analysis by logical means, before 1900 [42]. The second part, "Direct Proofs", presents in detail Hilbert's attempt to solve the consistency problem for elementary arithmetic in 1904, Poincaré's critical objections, and the impact of *Principia Mathematica* in Göttingen. "Proof Theoretic Strategies" is the title of the third part; here finitist mathematics moves to center stage, and the methodological perspective for Hilbert's proof theory is described. An analysis of the informal ideas underlying this approach leads to the fourth part, "Accessible Domains", and serves as the motivating background for a programmatic formulation of *reductive structuralism*. My goal there is twofold, namely, to describe a global, integrating perspective of foundational work on the one hand, and to formulate some more local, focused problems for mathematical work on the other hand.[11]

[9] See [44], p. 136. The German text refers to principle **I** of mathematical thought: "In **I**. kommt das schöpferische Prinzip zum Ausdruck, das uns im freiesten Gebrauch zu immer neuen Begriffsbildungen berechtigt mit der einzigen Beschränkung der Vermeidung eines Widerspruchs."

[10] See [7], p. 17. There Bernays locates first, in a most perspicuous way, the philosophical questions concerning mathematics. "Diese Fragen philosophischen Charakters haben eine besondere Dringlichkeit erhalten seit der Wandlung, welche die methodische Einstellung der Mathematik gegen Ende des 19. Jahrhunderts erfuhr." Then he describes the characteristic features of this transformation, namely, the advance of set theory, the emergence of existential axiomatics, and the forging of close connections between logic and mathematics.

[11] The considerations in this paper have profited from critical reactions to a number of earlier presentations, namely, at the Boolos Conference at Notre Dame (April 16, 1998), the Steinfest at the University of Chicago (May 23, 1999), the Hilbert Workshop at the Sorbonne (May 26, 2000), and the Annual Meeting of the Association for Symbolic Logic in Urbana (June 3, 2000). Special thanks for helpful criticism go to Bernd Buldt, Michael Detlefsen, Jacques Dubucs, Sol Feferman, Carl Posy, Howard Stein, and Bill Tait.

2 Logical Models

Hilbert attempted to secure analysis from contradictions at the close of the 19th century. His formulation of a theory for real numbers in 1899 (unpublished) was inspired by Dedekind's and is distinctly modern. Recall that Kronecker, a mere decade earlier, had still been trying to banish the general notion of irrational number from mathematics; and Hilbert's lecture notes from the period between (1894 *Quadratur des Kreises*; 1897 *Zahlbegriff und Quadratur des Kreises*; 1899 *Zahlbegriff und Quadratur des Kreises*; unpublished lecture notes) show, how difficult it was for him to obtain a proper perspective on the notion of number (*Zahlbegriff*). In the end, Hilbert associated all the central foundational issues with the *axiomatic method*. To proceed axiomatically means for Hilbert to think with consciousness, but also with critical awareness. The method allows the rigorous investigation of independence and completeness issues, and it is needed for securing, completely and logically, the content of our knowledge.[12] Already Dedekind had most explicitly aimed at grounding - by logic - our arithmetical knowledge!

2.1 Existence

A rather direct interpretation of the essay *Was sind und was sollen die Zahlen* and of the later explanatory letter to Keferstein shows that Dedekind strove to give a consistency proof relative to a logic that allowed the construction of models, here, for simply infinite systems. In other words, a logical proof of the existence of such a system was to secure that the very notion did not contain an "internal contradiction". Dedekind wrote to Keferstein:

> After the essential nature of the simply infinite system, whose abstract type is the number sequence N, had been recognized in my analysis . . ., the question arose: does such a system *exist* at all in the realm of our ideas? Without a logical proof of existence, it would always remain doubtful whether the notion of such a system might not perhaps contain internal contradictions. Hence the need for such a proof (articles 66 and 72 of my essay).[13]

These observations can be extended in a natural way to cuts and complete ordered fields, treated in the earlier essay *Stetigkeit und irrationale Zahlen* [23]. Dedekind viewed his broad methodological considerations not as specific for the foundational context of these essays, but rather as paradigmatic for the sound introduction of axiomatically characterized notions.[14]

In his proof of the existence of a simply infinite system, Dedekind had used however the (in)famous "system of all objects of my thinking". Hilbert learned

---

[12] The full German text is: "Trotz des hohen pädagogischen Wertes der genetischen Methode verdient doch zur endgültigen Darstellung und völligen logischen Sicherung des Inhaltes unserer Erkenntnis die axiomatische Methode den Vorzug."

[13] See [27], p. 101. The essay Dedekind refers to is obviously [25].

[14] Cf. See [26], pp. 268–69, in particular the long footnote on p. 269.

from Cantor as early as 1897 that this gave rise to a contradiction and, thus, undermined the logical basis of Dedekind's essay. The very title of Hilbert's historically sweeping lectures from 1894, *Quadratur des Kreises*, was consequently expanded in 1897 to *Zahlbegriff und Quadratur des Kreises*. Hilbert emphasized there the importance of the "fixation of the (real) number concept", and he defined the reals as fundamental sequences taking for granted the natural numbers. Two years later, when finishing *Grundlagen der Geometrie*, Hilbert formulated axioms for the reals including the completeness axiom in yet another version of these lectures on the quadrature of the circle. His paper *Über den Zahlbegriff* was completed for publication on October 12, 1899 and summarized these early investigations; Hilbert had presented it already on September 19, 1899 to the Munich meeting of the German Association of Mathematicians.[15]

A neglected, but most significant link to Dedekind should be noted. Hilbert followed Dedekind in formulating the central axiomatic conditions for real numbers, as well as in setting up the very framework by assuming the existence of a system of things satisfying the conditions, "We think a system of things..." (Wir denken ein System von Dingen...). He proceeded methodologically in exactly the same way for *Grundlagen der Geometrie*, where the existential framework for the axioms of geometry is introduced by "We think three different systems of things..." (Wir denken drei verschiedene Systeme von Dingen...). Thus, as in Dedekind's case, there is an explicit existential assumption that has to be secured or discharged in some way. To emphasize this crucial aspect of Hilbert's method, both Hilbert and Bernays called it *existential axiomatics* (existentiale Axiomatik). In the case of geometry Hilbert discharged the existential assumption for (parts of) the axiom system by appropriate analytic models. But how could the problem be addressed for analysis? How could the existence of the system of reals be secured? - In his answer, Hilbert referred to Cantor's distinction between consistent and inconsistent multiplicities that presented the former as the proper objects of set theory. Hilbert was critical of this distinction. His critical attitude, only implicit here, was made explicit in the Heidelberg talk of 1904 where he claimed that Cantor's conception "leaves latitude for subjective judgement and therefore affords no objective certainty".

2.2 A Partial Syntactic Turn

Dedekind had given a logical existence proof of a simply infinite system in order to guarantee that the very notion of such a system does not contain "internal contradictions". Hilbert recast consistency as a syntactic property of axiom systems, demanding that no contradiction be provable from the axioms in a finite number of steps.[16] That allowed him to attack the problem of arriving at a consistent

[15] In Peckhaus [65] the reader will find an informative, complementary discussion on pp. 29–33.

[16] He had done so also in section 9 of *Grundlagen der Geometrie*, but established consistency there semantically – by an "inductive argument" on pp. 19/20; cf. Note 12 of my (1999). Notice that Hilbert did not specify in either of these works the character of the "steps".

multiplicity of real numbers in *Über den Zahlbegriff* and his Paris lecture from a new viewpoint: the existence of sets is to be guaranteed by *consistency proofs* for appropriate axiom systems. Hilbert had shifted from consistent multiplicities to consistent theories[17] and suggested to give objective content to Cantor's notion through consistency proofs for theories. In the Paris lecture he thought that a *direct proof* should be possible:

> I am convinced that it must be possible to find a direct proof for the consistency of the arithmetical axioms [proposed in "Über den Zahlbegriff" for the reals, WS] by means of a careful study and suitable modification of the known methods of reasoning in the theory of irrational numbers.[18]

It is quite obscure from the published papers I referred to what would constitute a direct proof. A reasoned, though by no means unproblematic conjecture can be based on earlier lecture notes. Hilbert thought that the construction of the reals, and also of the natural numbers, could be given directly and be exploited as a blueprint for a Dedekindian consistency proof. This seems to be supported also by the (one-sidedly preserved) correspondence with Cantor at the time of the Munich and Paris talks. Cantor insisted in his letters on two points: (i) the consistency even of finite multiplicities (i.e., the existence of finite sets) has to be postulated, and (ii) Dedekind's considerations are fundamentally flawed. I conjecture Hilbert believed, despite the Cantorian admonitions, that Dedekind's logicism with suitable restrictions might after all provide the means for a principled consistency proof.[19]

Hilbert changed his views dramatically after Zermelo and Russell discovered their elementary contradiction, a contradiction that had according to his own testimony "a catastrophic effect in the mathematical world".[20] It had undoubtedly a catastrophic effect on Hilbert himself: Bernays reported to Constance Reid that Hilbert believed at the time, even if only for a very brief period, that Kronecker might have been right in demanding a radical restriction of mathematical notions and methods. In lecture notes from the summer of 1904, just before the Heidelberg talk in August of that year, one finds these illuminating and revealing remarks on Dedekind and Kronecker:

---

[17] Cf. See [15], p. 46; footnote 11 makes an explicit terminological recommendation w.r.t. "Konsistenz": "Es mag hier angeregt sein, diesen von Cantor speziell in bezug auf Mengenbildungen gebrauchten Ausdruck allgemein mit Bezug auf irgendwelche theoretischen Ansätze zu verwenden."

[18] See [43], p. 1104. This is re-emphasized in Bernays [10], p. 198–9: "Zur Durchführung des Nachweises gedachte Hilbert mit einer geeigneten Modifikation der in der Theorie der reellen Zahlen angewandten Methoden auszukommen."

[19] Cantor attended the Munich meeting and met with Hilbert. Cantor's views are carefully presented in his letter to Hilbert that was written on January 27, 1900. (The letter is contained in the Hilbert Nachla in Göttingen; Cod. 54 (18).)

[20] See [47], p. 169. "Insbesondere war es ein von Zermelo und Russell gefundener Widerspruch, dessen Bekanntwerden in der mathematischen Welt geradezu von katastrophaler Wirkung war." Some indication of related activities in Göttingen from1902 through 1904 is found in Peckhaus [65], p. 57.

> He [Dedekind] arrived at the opinion that the standpoint of viewing the integers as obvious cannot be sustained; he recognized that the difficulties Kronecker saw in the definition of irrationals arise already for integers; furthermore, if they are removed here, they disappear there. This work [*Was sind und was sollen die Zahlen*, WS] was epochal, but it did not yet provide something definitive, certain difficulties remain. These difficulties are connected, as with the definition of the irrationals, above all to the concept of the infinite; . . .[21]

These difficulties were plainly stated at the very beginning of the Heidelberg lecture, where Hilbert described in detail alternative foundational approaches and remarked about Dedekind:

> R. Dedekind clearly recognized the mathematical difficulties encountered when a foundation is sought for the notion of number; for the first time he offered a construction of the theory of integers, and in fact an extremely sagacious one. However, I would call his method *transcendental* insofar as in proving the existence of the infinite he follows a method that, though its fundamental idea is used also by philosophers, I cannot recognize as practicable or secure because it employs the notion of the totality of all objects, which involves an unavoidable contradiction.[22]

What could be done? – Hilbert shifted, first of all, his efforts from the theory of real numbers to that of integers; he proposed, secondly, to give a genuinely direct proof of the existence of “the smallest infinite”, and that was to be done by establishing the consistency of an axiom system that reflected Dedekind’s conditions for a simply infinite system.

3 Direct Proofs

The elementary Zermelo-Russell paradox had convinced Hilbert, as we saw, that there *was* a problem with his earlier considerations and that difficulties had to be faced at a more fundamental level. In the Heidelberg Lecture, Hilbert reasserted most strongly his view that the problems for the reals are resolved once matters are resolved for the natural numbers.

> The existence of the totality of real numbers can be demonstrated in a way similar to that in which the existence of the smallest infinite can be proved; in fact, the axioms for real numbers as I have set them up . . . can be expressed by precisely such formulas as the axioms hitherto assumed. . . . the axioms for the totality of real numbers do not differ qualitatively in any respect from, say, the axioms necessary for the definition of the integers. In the recognition of this fact lies, I believe, the real refutation of the conception of arithmetic associated with L. Kronecker . . .[23]

---

[21] (Hilbert 1904, unpublished), p. 166. “Er [Dedekind] drang zu der Ansicht durch, dass der Standpunkt mit der Selbstverständlichkeit der ganzen Zahlen nicht aufrecht zu erhalten ist; er erkannte, dass die Schwierigkeiten, die Kronecker bei der Definition der irrationalen Zahlen sah, schon bei den ganzen Zahlen auftreten und dass, wenn sie hier beseitigt sind, sie auch dort wegfallen. Diese Arbeit [*Was sind und was sollen die Zahlen*, WS] war epochemachend, aber sie lieferte doch noch nichts definitives, es bleiben gewisse Schwierigkeiten übrig. Diese bestehen hier, wie bei der Definition der irrationalen Zahlen, vor allem im Begriff des Unendlichen; . . .”

[22] See [44], pp. 130–1.

[23] See [44], pp. 137–8.

Hilbert actually claimed, "In the same way we can show that the fundamental notions of Cantor's set theory, in particular Cantor's alephs, have a consistent existence." Let us come back to the more modest goal of establishing the existence of the smallest infinite.

3.1 Turning Further

In Section 2.2, I described the partial syntactic turn Hilbert had taken by recasting consistency as a syntactic notion. However, he had neither specified inference steps nor had he presented other than semantic arguments. That was remedied here at least in a broad programmatic sense: he developed logic and arithmetic simultaneously and inferred the consistency of the joint system from elementary syntactic observations. The methodological starting-point was formulated in this way:

> Arithmetic is often considered to be a part of logic, and the traditional fundamental logical notions are usually presupposed when it is a question of establishing a foundation for arithmetic. If we observe attentively, however, we realize that in the traditional exposition of the laws of logic certain fundamental arithmetic notions are already used, for example, the notion of set and, to some extent, also the notion of number. Thus we find ourselves turning in a circle, and that is why a partly simultaneous development of the laws of logic and of arithmetic is required if paradoxes are to be avoided.[24]

The theory Hilbert proposed consists of axioms for identity and Dedekind's requirements for a simply infinite system, except that induction is not explicitly formulated; in modern notation:

(1) $\mathrm{x} = \mathrm{x}$
(2) $\mathrm{x} = \mathrm{y}$ & $\mathrm{A(x)}$–> $\mathrm{A(y)}$
(3) $\mathrm{x}' = \mathrm{y}'$–> $\mathrm{x} = \mathrm{y}$
(4) $\mathrm{x}' \neq 1$

The rules, extracted from Hilbert's description of "consequence", are modus ponens and a substitution rule that allows the replacement of variables by arbitrary sign combinations. Other modes of logical inferences are mentioned later, but neither formally stated nor incorporated into the consistency proof. The idea of the consistency proof is this: formulate a property P and show by induction on derivations that all provable equations have P. The property Hilbert considers is homogeneity: an equation $\mathrm{a} = \mathrm{b}$ is called *homogeneous* if and only if a and b have the same number of symbol occurrences; it is easily seen that all equations derivable from axioms (1)–(3) are indeed homogeneous. But a contradiction can be obtained only by establishing an unnegated instance of (4) from (1)–(3). Such an instance is necessarily inhomogeneous and, consequently, not provable.

Hilbert saw his considerations as answering, for the first time, the earlier call for a direct proof. He commented:

> The considerations just sketched constitute the *first case* [my emphasis, WS] in which a direct proof of consistency has been successfully carried out for axioms, whereas the

---

[24] See [44], p. 131.

> method of a suitable specialization, or of the construction of examples, which is otherwise customary for such proofs - in geometry in particular - necessarily fails here.[25]

Hilbert had emphasized the need to develop logic and mathematics simultaneously, but the actual work had significant shortcomings: there is no calculus for sentential logic, there is no proper treatment of quantification, and induction is neither rigorously formulated nor incorporated into the argument. In sum, there *is* an important shift from "semantic" arguments to a "syntactic" one, but the set-up is utterly incomplete as a formal framework for arithmetic.

3.2 Critical Analysis

The presumed foundational import of Hilbert's talk was not left unchallenged. On account of the inductive character of the consistency proof Poincaré criticized Hilbert's considerations severely; this critique is well-known and absolutely to the point. Towards the end of Hilbert's paper there is a peculiar "uncertainty" that reveals underlying methodological problems; they too were pointed out by Poincaré.[26] It becomes very clear, how penetrating Poincaré's considerations were, when one reads in parallel the 1935 remarks on Hilbert's Heidelberg talk by Bernays: they give a précis of Poincaré's critique. We should keep in mind, however, that the existence of sets was the central issue and was to be guaranteed by the consistency of an appropriate axiom system, a viewpoint shared explicitly and strongly by Poincaré. "If therefore", Poincaré wrote, "we have a system of postulates, and if we can demonstrate that these postulates imply no contradiction, we shall have the right to consider them as representing the definition of one of the notions entering therein."[27] In any event, as to the critical aspect Bernays wrote:

> . . . the systematic standpoint of Hilbert's proof theory is not yet fully and clearly developed. Some places indicate that Hilbert wants to avoid the intuitive conception of number and replace it by its axiomatic introduction. Such a procedure would lead to a circle in the proof theoretic considerations.[28]

---

[25] See [44], p. 135.

[26] on pp. 1042–3 in (Ewald). This is part of Poincaré's review of contemporaneous investigations of the foundations of mathematics [70]–[72]. The critical, but also sympathetic discussion of Hilbert is mainly found in Poincaré [71], pp. 1038–46. - Brouwer pointed to Poincaré, when elaborating on the first insight in his *Intuitionistic reflections on formalism*, and suggested that this insight had been "strongly prepared" by him.

[27] See [70], p. 1026. Having discussed Mill's view of (mathematical) existence and calling it "inadmissible", Poincaré writes in the immediately preceding paragraph: "Mathematics is independent of the existence of material objects; in mathematics the word 'exist' can have only one meaning; it means free from contradiction. . . . in defining a thing, we affirm that the definition implies no contradiction."

[28] See [10], p. 200. "Auerdem ist auch der methodische Standpunkt der Hilbertschen Beweistheorie in dem Heidelberger Vortrag noch nicht zur vollen Deutlichkeit entwickelt. Einige Stellen deuten darauf hin, da Hilbert die anschauliche Zahlvorstellung vermeiden und durch die axiomatische

Bernays emphasized also that Hilbert had not articulated distinctions central to the later finitist program:

> Also, the viewpoint of restricting the contentual application of the forms of existential and general judgements is not yet put forth explicitly and completely.[29]

Hilbert's own views of his objective accomplishments were formulated in lectures from the summer term of 1905 [44].[30] These lectures contain additional technical details, but point out basic shortcomings as well; Hilbert bemoans the unsatisfactory state of logic, in particular, the state of quantification theory. Hilbert had a distinctive approach already in the Heidelberg lecture, clearly recognized and applauded by Poincaré, as Hilbert's "all" ranged only over the limited domain of combinations of thought-objects, not as Russell's over everything whatsoever; cf. p. 1040 of Poincaré [70].

3.3 Proper Formalisms

During the period from 1905 to 1917 Hilbert gave almost annually lectures on the foundations of mathematics, but these lectures did neither break new ground, nor did they return to a proof theoretic study;[31] another approach opened up, however. Around 1913 Hilbert started to become familiar with some of Russell's writings. The official lecture notes from the winter term 1914/15 contain brief remarks about type theory, and the notes from a student, serendipitously preserved in the Institute for Advanced Study at Princeton, more extended ones. There was even some correspondence between Hilbert and Russell, reported in Appendix B of Sieg [80]. A number of relevant talks on the foundations of mathematics were given in Göttingen during this period by Behmann, Bernstein, Hilbert, and Zermelo; most significantly, Hilbert directed Behmann's dissertation of 1918, *Die Antinomie der transfiniten Zahl und ihre Auflösung durch die Theorie von Russell und Whitehead*. A detailed description of this work is found in Mancosu [55], narrowing

Einführung des Zahlbegriffs ersetzen will. Ein solches Verfahren würde in den beweistheoretischen Überlegungen einen Zirkel ergeben."

[29] See [10], p. 200. "Auch wird der Gesichtspunkt der Beschränkung in der inhaltlichen Anwendung der Formen des existentialen und des allgemeinen Urteils noch nicht ausdrücklich und restlos zur Geltung gebracht."

[30] These lectures are discussed in detail by Peckhaus in [65, 66, 67, 68]; a broad philosophical perspective is also provided by Hallett in [37, 38].

[31] That is supported, in a general way, by Bernays [10]. However, in Bernays's description there is a peculiar "smoothing" of the developments between 1904/5 and 1917/22: Bernays does not mention that Hilbert gave lectures on the foundations of mathematics during that period; the 1917/8 lectures are not hinted at. Thus, he effectively creates the impression that the period is one of inactivity; e.g., on p. 200 one finds: "In diesem vorläufigen Stadium hat Hilbert seine Untersuchungen über die Grundlagen der Arithmetik für lange Zeit unterbrochen. Ihre Wiederaufnahme finden wir angekündigt in dem 1917 gehaltenen Vortrage 'Axiomatisches Denken'." This impression is reinforced by the footnote attached to the first sentence in this quote, where Bernays points to the work of others who pursued the research direction stimulated by Hilbert [44].

the real gap in our historical understanding of the details of the Russellian influence on Hilbert.[32] How strongly Russell influenced Hilbert has been clear from the notes for his course on *Set Theory* (summer term 1917) and his Zürich talk *Axiomatisches Denken* given on September 11, 1917; they reveal renewed logicist tendencies in Hilbert's work. Hilbert wrote in the essay on which his talk had been based:

> The examination of consistency is an unavoidable task; thus, it seems to be necessary to axiomatize logic itself and to *show that number theory as well as set theory are just parts of logic*.

If we try to achieve such a reduction to logic, Hilbert said at the very end of the set theory notes, "... we are facing one of the most difficult problems of mathematics".[33]

Russell and Whitehead provided not only the stimulus for this programmatic re-direction, but also powerful technical tools.[34] The latter were ingeniously adapted and mathematically analyzed in the winter term 1917/18 (Prinzipien der Mathematik, unpublished lecture note), when Hilbert offered lectures under the title *Prinzipien der Mathematik* with the assistance of Paul Bernays. Hilbert had finally a proper formalism for the development of mathematics: a language for capturing the logical form of informal statements and a calculus for representing the structure of logical arguments. The presentation is carried through with focus, elegance, and directness. The logical work is complemented by real metamathematical considerations; the latter are certainly inspired by the (perspective underlying the) work that had been done at the turn of the century on the foundations of geometry. These beautifully written, detailed notes include all the basic material that is contained in the 1928 book by Hilbert & Ackermann; thus, they are the real beginning of modern *mathematical* logic. For my purpose the main points can be summarized as follows: (i) there is a full development of (the syntax and semantics for) sentential, monadic, first-order logic, and ramified analysis, (ii) independence and completeness problems are formulated and partly solved, and (iii) theories are always presented with appropriate domains or, more precisely, many-sorted structures. The last point brings out in this setting the crucial aspect of existential axiomatics that had been so important in Hilbert's early investigations (cf. Section 2.1).

Absolutely no proof theoretic considerations are presented in these notes, though concistency is a real issue. The consistency of pure logic is examined, and both sentential and first order logic are semantically shown to be consistent, the latter

---

[32] The list of talks is found on pp. 304–5 of Mancosu [55]. Behmann is viewed by Mancosu as a central player and as the (indirect) source for some of Hilbert's views, i.e., as a conduit for Russellian views.

[33] That is not at all reflected in Bernays's presentation in [10]; the logicist tendency is suppressed and the "ungelöste Problematik" of the consistency of *Principia Mathematica* is emphasized immediately; cf. p. 201.

[34] How much direct continuity is there between *Principia Mathematica* [89]–[91] and *Prinzipien der Mathematik*? That remains an important question for detailed investigation.

by considering a one-element domain. A footnote warns the reader, however, not to overestimate the significance of this result for first-order logic, because "[i]t does not give us a guarantee that the system of provable formulas remains free of contradictions after the symbolic introduction of contentually correct assumptions".[35] After all, these assumptions may force the domain to be infinite.

At the beginning of 1920, having abandoned for good reasons the logicist route and responding in part to the contemporaneous investigations of Brouwer and Weyl, Hilbert and Bernays pursued a radically constructive redevelopment of arithmetic. This took up a recurring Kroneckerian theme in Hilbert's foundational reflections. However, it was realized quickly that this could not provide a foundation for classical forms of reasoning, as the law of excluded middle does not hold for constructively understood quantified statements. Having recognized this fact, Hilbert and Bernays mentioned Brouwer for the first time when closing with: "This consideration helps us to gain an understanding for the sense of the paradoxical claim, made recently by Brouwer, that for infinite systems the law of the excluded middle (the 'tertium non datur') loses its validity."[36]

To us it may seem as if Hilbert had available all the mathematical and logical means for the formulation of *the* program. Yet it took some more time before he had gained the appropriate methodological perspective, and before finitist mathematics and proof theory emerged in a programmatically coherent alignment.

4 Proof Theoretic Strategies

Hilbert had argued for a *theory of proofs* in his 1904 Heidelberg talk; he had mentioned it also in his 1917 Zürich talk, but without any programmatic direction. The suggestion was finally taken up again in the summer semester of 1920: the notes from that term contain a consistency proof for a restricted part of elementary number theory. Indeed, it is (almost exactly) the system of the Heidelberg Lecture.[37]

---

[35] (Hilbert 1917/18, unpublished), p. 156. "Man darf dieses Ergebnis in seiner Bedeutung nicht überschätzen. Wir haben ja damit noch keine Gewähr, dass bei der symbolischen Einführung von inhaltlich einwandfreien Voraussetzungen das System der beweisbaren Formeln widerspruchsfrei bleibt."

[36] Cf. Sieg[80], pp. 23–27, for more details on the attempted radical constructive development.

[37] The language of this fragment of arithmetic consists of variables a, b, . . ., non-logical constants 1, +, and all numerals; = and are the only relation symbols, and –> is the sole logical symbol. The axioms are:

(1) $1 = 1$

(2) $\mathrm{a} = \mathrm{b} - > \mathrm{a} + 1 = \mathrm{b} + 1$

(3) $\mathrm{a} + 1 = \mathrm{b} + 1 - > \mathrm{a} = \mathrm{b}$

(4) $\mathrm{a} = \mathrm{b} - > (\mathrm{a} = \mathrm{c} - > \mathrm{b} = \mathrm{c})$

(5) $\mathrm{a} + 1 \neq 1$

As to inference rules we have modus ponens and a substitution rule for numerals.

4.1 Turning Further, Ctd.

The syntactic turn in treating the consistency problem, from choosing a syntactic formulation of consistency to developing logic and arithmetic simultaneously, is pursued further in these notes. The description of the system of elementary arithmetic is given in a more coherent way and is evidently informed by the logical work of the prior years. Attention is paid to the mathematical means used in the proof theoretic arguments, but the formalism that is being investigated is semi-constructive; cf. the end of this subsection. The formalism is almost exactly that of the Heidelberg Lecture, but the argument for its consistency is quite different, mainly through the introduction and use of the notion "kürzbar":

> . . . If one considers a proof with respect to a particular concrete property it has, then it is possible that the removal of some formulae in this proof still leaves us with a proof that has that particular property. In this case we are going to say that the proof is *kürzbar* with respect to the given property.[38]

Hilbert establishes three lemmata: the first claims that a theorem can contain at most two occurrences of $->$, the second asserts that no statement of the form $(A-> B)-> C$ can be proved, and the third expresses that a formula $a = b$ is provable only if a and b are the same term. To recognize the distinctive character of the arguments, let me look at the proof of the first lemma. Hilbert proceeds indirectly and assumes that there is a theorem with at least three occurrences of $->$. Without loss of generality he further assumes that the theorem has a proof that is not "kürzbar" (w.r.t. the property of having an endformula with at least three occurrences of the $->$). The theorem cannot be an axiom, as axioms contain at most two occurrences of the $->$. Thus, it must have been obtained by modus ponens. The major premise of that inference contains at least one more occurrence of $->$ than its conclusion, i.e., the given theorem; we have consequently a contradiction to the "Nicht-Kürzbarkeit" of the given proof.

On its surface, Hilbert's new proof does not use the induction principle. It is structured in analogy to the standard proof of the fact that $\sqrt{2}$ is not rational. Hilbert frequently asserted, not only here, that consistency proofs should be of the same character as the proof of the irrationality of $\sqrt{2}$.[39] The analogy plays even on the

---

[38] (Hilbert 1920, unpublished), p. 38. "Betrachtet man einen Beweis in Hinsicht auf eine bestimmte, konkret aufweisbare Eigenschaft, welche er besitzt, so kann es sein, dass, nach Wegstreichung einiger Formeln in diesem Beweise noch immer ein Beweis (. . .) übrig bleibt, welcher auch noch jene Eigenschaft besitzt. In diesem Fall wollen wir sagen, dass der Beweis sich in bezug auf die betreffende Eigenschaft kürzen lässt."

[39] On pages 7a/8a of (Hilbert 1921/22, unpublished) one finds the remark: "Diese Aufgabe [to show that it is impossible to derive in a given calculus certain formulas like $1 \neq 1$] liegt grundsätzlich ebenso im Bereich der anschaulichen Betrachtung wie etwa die Aufgabe des Beweises, dass es unmöglich ist, zwei Zahlzeichen a, b zu finden, welche in der Beziehung $a^2 = 2b^2$ stehen. Hier soll gezeigt werden, dass sich nicht zwei Zahlzeichen von einer gewissen Beschaffenheit angeben lassen. Entsprechend kommt es für uns darauf an zu zeigen, dass sich nicht ein Beweis von einer bestimmten Beschaffenheit angeben lässt. Ein formalisierter Beweis ist aber, ebenso wie ein Zahlzeichen, ein konkreter und überblickbarer Gegenstand. Er is (wenigstens grundsätzlich)

double meaning of "kürzbar"; on the one hand, "kürzbar" means most directly, when applied to proofs, "can be shortened", but on the other hand it also applies to fractions and means then "(a common factor) can be canceled". Recall that the standard argument proceeds also indirectly, assuming that $\sqrt{2}$ is rational, i.e., equals p/q, q $\neq$ 0; without loss of generality it is then assumed further that p/q is not "kürzbar". In his 1922 publication, based on lectures he had given in the spring and summer of 1921 in Kopenhagen and Hamburg (and submitted for publication not before November of 1921), Hilbert makes explicit the strategic point of the modified argument:

> Poincaré's objection, that the principle of complete induction cannot be proved but by complete induction, has been refuted by my theory.[40]

Is this to be taken in the strong sense that induction is not used at all? Or is it to be understood, perhaps, just in the weaker sense that a special procedure is being used – a procedure based on the construction and deconstruction of numerals and that, by its very nature, is different from the induction principle?[41] From a

---

von Anfang bis Ende mitteilbar. Auch die verlangte Eigenschaft der Endformel, z.B. dass sie "1 $\neq$ 1" lautet, ist eine konkret feststellbare Eigenschaft des Beweises." In the Lecture Notes from the following year this view is expressed again as follows (p. 33): "Hier kommt es zur Geltung, dass die Beweise, wenn sie auch inhaltlich sich im Transfiniten bewegen, doch, als Gegenstände genommen und formalisiert, von finiter Struktur sind. Aus diesem Grunde ist die Behauptung, dass aus bestimmten Aussagen nicht zwei Formeln A, $\neg$A bewiesen werden können, methodisch gleichzustellen mit inhaltlichen Behauptungen der anschaulichen Zahlentheorie, wie z.B. der, dass man nicht zwei Zahlzeichen a, b finden kann, für welche $a^2 = 2b^2$ gilt." That is also asserted in publications, for example, in *On the Infinite*, p. 383 of (van Heijenoort).

In Bernays, [10], p. 76, one finds this remark about the character of consistency proofs: "Diese Unmöglichkeitsbehauptung, um deren Beweis es sich hier handelt, hat die gleiche Struktur wie z.B. die Behauptung, da es unmöglich ist, die Gleichung $a^2 = 2b^2$ durch zwei ganze Zahlen a und b zu erfüllen."

40 See [45], p. 161. "Sein [Poincarés] Einwand, dieses Prinzip [der vollständigen Induktion] könnte nicht anders als selbst durch vollständige Induktion bewiesen werden, ist durch meine Theorie widerlegt."

41 Bernays, in his contemporaneous paper (1922) [4], acknowledges explicitly that a form of induction has to be used; indeed, the writing of Bernays [4] preceded the 1921/22 Lectures, where induction is discussed (in particular on p. 57) and used for the development of finitist arithmetic. Reflecting on the proof of the commutativity of addition, Hilbert and Bernays write (on pp. 56–7): "Bei diesem Beweis wenden wir eine Art von vollständiger Induktion an, die aber auch, in der Form, wie sie hier gebraucht wird, ganz dem Standpunkt unserer anschaulichen Betrachtungsweise entspricht. Das Beweisverfahren kommt auf einen Abbau der Zahlzeichen hinaus, d.h. wir benutzen die Tatsache, dass die Zahlzeichen, ebenso wie sie durch Zusammensetzung von 1 und + aufgebaut sind, sich auch umgekehrt durch Wegnahme von 1 und + abbauen lassen müssen." Bernays discusses matters in a very similar manner in his (1935) [10], p. 203: "Was ferner die methodische Einstellung betrifft, welche Hilbert seiner Beweistheorie zugrunde legt und welche er an Hand der anschaulichen Zahlentheorie erläutert, so liegt darin – ungeachtet der Stellungnahme Hilberts gegen Kronecker – eine weitgehende Annäherung an den Standpunkt Kroneckers vor. Eine solche besteht insbesondere in der Anwendung des anschaulichen Begriffes der Ziffer und ferner darin, da die anschauliche Form der vollständigen Induktion, d.h. die Schluweise, welche sich auf die anschauliche Vorstellung von dem 'Aufbau' der Ziffern gründet, als einsichtig und keiner weiteren Zurückführung bedürftig anerkannt wird. Indem Hilbert sich zur Annahme dieser

mathematical point of view Hilbert used the least number principle in an elementary form, namely, applied to a purely existential statement. The work in this first of Hilbert's foundational articles of the twenties is evidently transitional. It does have a major problem in not recognizing clearly necessary metamathematical means, but also in not fixing appropriately the very logic of the formal system to be investigated: Hilbert tried to keep it constructive by using, for example, negation only in a restricted way.[42]

4.2 Principled Formulation

Proof theoretic considerations were pursued with novel metamathematical means and with a principled foundational perspective in the lectures from the winter term 1921/22. For the first time, Hilbert and Bernays used the terms *finitist mathematics* and *Hilbert's proof theory* and made explicit the domain of mathematical (finitist) objects appealed to in proof theoretic investigations. They pointed out:

> We have to extend the domain of objects to be considered; i.e., we have to apply our intuitive considerations also to figures that are not number signs. Thus, we have good reason to distance ourselves from the earlier dominant principle according to which each theorem of pure mathematics is in the end a statement concerning integers.

With a jibe at such distinguished mathematicians as Dirichlet and Dedekind, they continued, "This principle was viewed as expressing a fundamental methodological insight, but it has to be given up as a prejudice."[43] After all, formulas and proofs of formal theories are the direct object of proof theoretic investigation, and appropriate definition and proof principles (analogous to those for numbers) have to be used.

Hilbert proved the consistency of a fragment of number theory with *full classical sentential logic* and free variable statements. The very elaborate and detailed proof given in the Notes was sketched in Hilbert's Leipzig talk of September 22, 1922 [45].[44] A treatment of quantifiers is indicated there, and genuine transformations of

---

methodischen Voraussetzung entschlo, wurde auch der Grund der Einwendungen behoben, welche seinerzeit Poincaré gegen Hilberts Unternehmen in dem Heidelberger Vortrag gerichtet hatte."

[42] For a more detailed discussion of this "transitional stage", see [80], pp. 26–7 and Appendix A. – The "proper" response to Poincaré is formulated in Hilbert's second Hamburg Talk, [48], p. 473; cf. also the remarks by Bernays quoted at the end of the preceding note.

[43] Hilbert 1921/22, p. 4a. "Wir müssen den Bereich der betrachteten Gegenstände erweitern, d.h. wir müssen unsere anschaulichen Überlegungen auch auf andere Figuren als Zahlzeichen anwenden. Wir sehen uns somit veranlasst, von dem früher herrschenden Grundsatz abzugehen, wonach jeder Satz der reinen Mathematik letzten Endes in einer Aussage über ganze Zahlen bestehen sollte. Dieses Prinzip, in welchem man eine grundlegende methodische Erkenntnis erblickt hat, müssen wir jetzt als Vorurteil preisgeben."

[44] The text of the lecture was submitted to Mathematische Annalen on September 29, 1922, and published in 1923 as *Die logischen Grundlagen der Mathematik*.

formal proofs are used to carry out the argument. Equally striking is the underlying idea that expresses in a novel, precise way the 19th century methodological maxim that elementary statements should be provable by elementary means. *Elementary* statements are those formulas, also called *numeric*, that are built up solely from $=$, $\neq$, numerals, and sentential logical connectives.

The central steps of the proof theoretic argument are described easily: (i) formal proofs with a numeric endformula are transformed into configurations that are not necessarily proofs, but consist only of numeric formulas; (ii) formulas in these configurations are all effectively brought into normal forms; (iii) the resulting normal form statements are all recognized to be "true". Given a formal proof of $0 \neq 0$, the transformations leave the endformula unchanged. From (iii) and the fact that $0 \neq 0$ is not true, it follows that $0 \neq 0$ is not provable. Clearly, these considerations are preliminary in the sense that they concern a theory that is part of finitist mathematics and thus need not be secured by a consistency proof. The next step is crucial w.r.t. the real issue of securing parts of mathematics that properly extend finitist mathematics.

Hilbert treats quantifiers with the τ-function, the dual of the later ε-operator; τ associates with every predicate A(a) a particular object $\tau_a(A(a))$ or simply τA. The *transfinite axiom* A(τA)-> A(a) expresses, according to Hilbert, "if a predicate A holds for the object τA, then it holds for all objects a". The τ-operator allows the definition of the quantifiers:

$$(\forall a)\ A(a) <\!\!-\!\!> A(\tau A),$$
$$(\exists a)\ A(a) <\!\!-\!\!> A(\tau(\neg A)).$$

Hilbert extends the proof theoretic considerations to the "first and simplest case" of going beyond the finitist system. The technique used will become the ε-substitution method, allowing the elimination of quantifiers from proofs of quantifier-free statements. Thus, finitist proof theory is given not only its principled formulation, but also its guiding idea (reflection principle[45]) and a dual version of its technical tool (ε-calculus). Bernays writes in 1935 [10]: "With the presentation of proof theory as given in the Leipzig talk the principled form of its structure had been reached."[46] Ackermann's thesis, published as [1], is a direct continuation of Hilbert's paper.

---

[45] Already in the Lecture Notes of 1921/22 we find, on p. 4a, this remark, after a discussion of the "incorrect application of the law of the excluded middle": "Wir sehen also, dass für den Zweck einer strengen Begründung der Mathematik die üblichen Schlussweisen der Analysis in der Tat nicht als logisch selbstverständlich übernommen werden dürfen. Vielmehr ist es gerade erst die Aufgabe für die Begründung, zu erkennen, warum und in wieweit die Anwendung der transfiniten Schlussweisen, so wie sie in der Analysis und in der (axiomatisch begründeten) Mengenlehre geschieht, stets richtige Resultate liefert."

[46] See [10], p. 204. "Mit der Gestaltung der Beweistheorie, die uns in dem Leipziger Vortrag entgegentritt, war die grundsätzliche Form ihrer Anlage erreicht." Bernays gives on that page also a summary of the crucial features of the proof.

4.3 Uniform Projection

Instead of pursuing the all-too-well-known sequence that starts with Hilbert and Bernays, goes through Ackermann, von Neumann, Herbrand, and then ends with Gödel, I turn to the question: What is the informal idea underlying the proof theoretic work, including the very idea of formalizing mathematical theories? - An answer to this question is found most directly in papers by Bernays from 1922 [4] and in the related lecture notes from 1921/22 (*Grundlagen der Mathematik*, unpublished lecture note) and 1922/23.[47] As we saw, Hilbert's *existential axiomatics* assumed always a system of objects satisfying the axiomatic conditions, and Bernays remarked:

> In the assumption of such a system with particular structural properties lies something so-to-speak transcendental for mathematics, and the question arises which principled position with respect to it should be taken.

An intuitive grasp of the completed sequence of natural numbers, for example, or of the manifold of real numbers should not be excluded outright. However, taking into account tendencies in the exact sciences, Bernays suggested a different strategic direction, namely, to try "whether it is not possible to give a foundation to these transcendental assumptions in such a way that only primitive intuitive knowledge is used." This suggestion is supplemented by a wonderful image of how to exploit the formalizability of axiomatic theories for this goal: their formalization serves to *project* the associated structures uniformly into the proper mathematical, finitist domain. Even fifty years later Bernays used that image and emphasized the epistemological significance of such projections:

> In taking the deductive structure of a formalized theory ... as an object of investigation the [contentual] theory is projected as it were into the number theoretic domain. The number theoretic structure thus obtained is in general essentially different from the structure intended by the [contentual] theory. But it [the number theoretic structure] can serve to recognize the consistency of the theory from a standpoint that is more elementary than the assumption of the intended structure.[48]

---

[47] The 1921/22 Lectures, p. 7a, emphasize the methodological point of formalization (and its relation to *existential axiomatics*) as follows:

"Diesen Formalismus können wir nun zum Gegenstand einer anschaulichen Betrachtung machen, und damit eröffnet sich uns die Möglichkeit einer strengen Begründung der Mathematik.

Denn das Problem der Widerspruchsfreiheit, welches ja die grundsätzlichen Schwierigkeiten bot, erhält durch den neuen Standpunkt eine ganz konkrete Fassung. Es handelt sich nicht mehr darum, ein System von unendlich vielen Dingen mit bestimmten Verknüpfungseigenschaften (eine stetige Mannigfaltigkeit von gewisser Art) als logisch möglich zu erweisen, sondern es kommt nur darauf an einzusehen, dass es unmöglich ist, aus den (in Formeln aufgeschriebenen) Axiomen nach den Regeln des logischen Kalküls gewisse Formeln wie z.B. $1 \neq 1$ abzuleiten."

[48] See [14], p. 186. The same image is used in the almost contemporaneous correspondence with Gödel; Bernays, after describing how intuitionism considers proofs as proper objects of mathematics, remarks in his letter of March 16, 1972: "Gewiss macht auch die Hilbert'sche Metamathematik die mathematischen Beweise zum Gegenstand, aber doch nur, nachdem sie diese durch die Formalisierung gleichsam in die mathematische Gegenständlichkeit projiziert hat."

The reader may consider this image as merely playful or as genuinely helpful. I choose the latter view, because the substantive point can be recast, and was recast by Bernays in his (1930) [7], as an explication of Hilbert's existential axiomatics that reveals a thoroughly structuralist perspective. "Structuralist" is here to be taken in the modern philosophical sense as described so masterfully in Parsons [63] and discussed extensively in Shapiro [74]. Parsons states there that views of this kind can be traced back to the end of the 19th century, but attributes clear general statements only to Bernays in 1950 [12] and to Quine somewhat later.[49] However, Bernays points already in his (1930) [7] to a characteristic aspect of the *newer mathematics* and describes the subject repeatedly as the study of structures (e.g., on p. 32). He presents concisely the standard account of if-then-ism or deductivism, and raises – as the starting point of his systematic philosophical investigations – the vacuity issue for that account. He takes this problematic as arising from two moments of modern axiomatics, namely, (i) the purely hypothetical connection between axioms and theorems, abstracting from the content and truth of the axioms, and (ii) the existential formulation of mathematical theories, assuming a given and from the very beginning fixed system of things and relations pertaining to them.[50] Let me present Bernays's discussion (on pp. 20–21) of the central point in greater detail.

Given the perspective on modern axiomatics I just sketched, the axioms and theorems of an axiomatic theory are statements that concern the relations occurring in them, and the relations pertain to the things of an assumed system. The knowledge provided by a proof of a theorem (Lehrsatz) L from axioms $A_1, \ldots, A_k$ consists in the determination (Feststellung) that, if the statements $A_1, \ldots, A_k$ hold for the

---

[49] Parsons refers to [12] and states w.r.t. Quine: "Quine is generally most explicit when speaking of natural numbers. For a very explicit general statement, however, see *Ontological Relativity and Other Essays*, (Columbia University Press, New York, 1969), pp. 43–45."

[50] The first moment is beautifully formulated on pp. 3–4 of the 1921/22 Lecture Notes (and takes up the theme of the Paris Lecture quoted above): "Auf diese Weise bildete sich die Einsicht heraus, dass das Wesentliche an der axiomatischen Methode nicht in der Gewinnung einer absoluten Sicherheit besteht, die auf logischem Wege von den Axiomen auf die Lehrsätze übertragen wird, sondern darin, dass die Untersuchung der logischen Zusammenhänge von der Frage der sachlichen Wahrheit abgesondert wird.
Unter diesem Gesichtspunkt stellt sich die Methode des axiomatischen Aufbaues einer Theorie dar als ein Verfahren der Abbildung eines Wissensgebietes auf ein Fachwerk von Begriffen, welche so geschieht, dass den Gegenständen des Wissensgebietes die Begriffe und den Aussagen über die Gegenstände die logischen Beziehungen zwischen den Begriffen entsprechen.
Durch diese Abbildung wird die Untersuchung von der konkreten Wirklichkeit ganz losgelöst. Die Theorie hat mit den realen Objekten und mit dem anschaulichen Inhalt der Erkenntnis gar nichts mehr zu tun; sie ist ein reines Gedankengebilde, von dem man nicht sagen kann, dass es wahr oder falsch ist. Dennoch hat dieses Fachwerk von Begriffen eine Bedeutung für die Erkenntnis der Wirklichkeit, weil es eine mögliche Form von wirklichen Zusammenhängen darstellt. Die Aufgabe der Mathematik ist es, solche Begriffsfachwerke logisch zu entwickeln, sei es, dass man von der Erfahrung her oder durch systematische Spekulation auf sie geführt wird.
Hier erhebt sich nun die Frage, ob denn jedes beliebige Fachwerk ein Abbild wirklicher Zusammenhänge sein kann. Eine Bedingung ist dafür jedenfalls notwendig: Die Sätze der Theorie dürfen einander nicht widersprechen, das heisst, die Theorie muss in sich möglich sein, somit entsteht das *Problem der Widerspruchsfreiheit*."

relations, then the statement L also holds for these relations. Here we have, as Bernays puts it, a general theorem on relations, i.e., a theorem of pure logic: the results of an axiomatic theory present themselves as theorems of logic. However, these theorems have significance only if the axiomatic conditions can be satisfied at all:

> If such a satisfying structure is unthinkable, i.e., logically impossible, then the axiom system does not lead to any theory at all, and the only logically meaningful statement concerning the system [of axioms, WS] is thus the determination (Feststellung) of the contradiction following from the axioms. For this reason there is for every axiomatic theory the requirement of a proof of the *satisfiability*, i.e., the *consistency* of its axioms.[51]

Bernays observes further that such proofs are usually given by providing arithmetical models, unless one can get by with the construction of finite ones. Thus, Bernays has retraced Hilbert's motivation for a consistency proof or, perhaps better, isolated the methodological core of his considerations. What is surprising after Herbrand's and Gödel's dissertations is, what seems to be, an unapologetic identification of satisfiability and consistency. In an unpublished note (found in the Appendix), Bernays describes this connection properly and in harmony with the principles guiding proof theoretic investigations. In any event, Bernays formulates here first a position of (what Parsons calls) *eliminative structuralism* in a concise way. That position is complemented by principled, philosophical reflections and programmatic, mathematical efforts to obtain finitist consistency proofs. It is this additional reflective perspective that allows us to see Hilbert and Bernays's structuralism as being of the non-eliminative variety; cf. the beginning of Section 5.2.

We saw how Hilbert and Bernays tried to exploit the special epistemological status of finitist mathematics for consistency proofs. After the discovery of Gödel's Second Incompleteness Theorem, however, the fundamental status of finitist mathematics had to be given up, and finitist considerations had to be expanded by considerations in stronger theories. To avoid a threatening vicious circle, as in Hilbert [44], these stronger theories had to be constructively motivated. The first consistency proof satisfying such an informal demand was obtained independently by Gödel and Gentzen and established the consistency of classical arithmetic relative to its intuitionistic version: the latter theory was indeed based on an extended constructive viewpoint, ironically, the intuitionistic one. Bernays in his (1954) [13] called this *sharpened axiomatics* (verschärfte Axiomatik) and formulated as a minimal requirement that "the objects [making up the intended model of the theory, WS] are not taken from a domain that is thought as being already given, but are rather

---

[51] See [7], p. 21. "Ist eine solche Erfüllung undenkbar, d.h. logisch unmöglich, so führt das Axiomensystem zu gar keiner Theorie, und die einzige logisch belangvolle Aussage über das System [von Axiomen, WS] ist dann die Feststellung des aus den Axiomen sich ergebenden Widerspruchs. Aus diesem Grunde besteht für jede axiomatische Theorie die Erforderlichkeit eines Nachweises der *Erfüllbarkeit*, d.h. der *Widerspruchsfreiheit* ihrer Axiome." – It should be evident from my earlier discussion that Dedekind saw both of these moments very clearly and, consequently, formulated the consistency problem most appropriately and sharply. He tried to resolve it by model theoretic considerations within logic.

constituted by generative processes".[52] There is no indication in Bernays's (1954) [13] or in his later writings, what kind of generative processes should be considered, and why that particular feature of domains should play a distinctive, foundational role. These two issues are at the center of the considerations in the next section.

5 Accessible Domains

When Gödel considered Platonism still as a doctrine "which cannot justify any critical mind and which does not even produce the conviction that they [the axioms of set theory, WS] are consistent", he analyzed also different layers of constructive mathematics in most informative ways. The lowest layer, identified with finitist mathematics, had one important characteristic:

> The application of the notion "all" or "any" is to be restricted to those infinite totalities for which we can give a finite procedure for generating all their elements (as we can, e.g., for the totality of integers by the process of forming the next greater integer and as we cannot, e.g., for the totality of all properties of integers).[53]

Directly associated with the procedure for generating the integers are the principles of proof by induction and definition by recursion. There is no further analysis of this direct association; the principles are simply taken to have a high degree of evidence. Can one go beyond such a brief, purely descriptive account and, perhaps, extend the considerations to other classes of mathematical objects?

5.1 Finitist Objects and Processes

Gödel considered the totality of integers as just one example of totalities whose elements are generated by a finite procedure. That a greater class of such totalities

---

[52] See [13], pp. 11–2. "Die Mindest-Anforderung an eine verschärfte Axiomatik ist die, dass die Gegenstände nicht einem als vorgängig gedachten Bereich entnommen werden, sondern durch Erzeugungsprozesse konstituiert werden." Bernays continues with a methodologically important remark: "Es kann aber dabei die Meinung sein, dass durch diese Erzeugungsprozesse der Umkreis der Gegenstände determiniert ist; bei dieser Auffassung erhält das *tertium non datur* seine Motivierung. In der Tat kann Offenheit eines Bereiches in zweierlei Sinn verstanden werden, einmal nur so, dass die Konstruktionsprozesse über jeden einzelnen Gegenstand hinausführen, und andererseits in dem Sinne, dass der resultierende Bereich überhaupt nicht eine mathematisch bestimmte Mannigfaltigkeit darstellt. Je nachdem die Zahlenreihe in dem erstgenannten oder in dem zweiten Sinne aufgefasst wird, hat man die Anerkennung des *tertium non datur* in bezug auf die Zahlen oder den intuitionistischen Standpunkt. Bei dem finiten Standpunkt kommt noch die Anforderung hinzu, dass die Überlegungen an Hand der Betrachtung von endlichen Konfigurationen verlaufen, somit insbesondere Annahmen in der Form allgemeiner Sätze ausgeschlossen werden."

[53] See [33], p. 51. The reflections in this paper are continued most directly in Gödel's *Lecture at Zilsel's*, [34]; the latter notes contain in particular a detailed analysis of Gentzen's first consistency proof for elementary number theory.

has directly associated principles had been emphasized already by Poincaré. After a discussion of the induction principle for natural numbers in his (1905), he remarked:

> I did not mean to say, as has been supposed, that all mathematical reasonings can be reduced to an application of this principle. Examining these reasonings closely, we should see applied there many analogous principles, presenting the same essential characteristics. In this category of principles, that of complete induction is only the simplest of all and this is why I have chosen it as a type. (p. 1025)

Modern expositions and critical examinations of Hilbert's considerations, e.g., those of Parsons and Tait [84, 85], focus on natural numbers. As a matter of fact, so did Bernays in his (1930) [7], but he viewed the case of numbers as paradigmatic and embedded it into broader reflections on the nature of mathematical knowledge; the latter was to be captured in a principled way, independently of the current inventory of mathematical disciplines. Bernays viewed a *certain kind of abstraction* as distinctive for the nature of mathematical thought:

> This abstraction may be called formal or mathematical abstraction. It consists in emphasizing the structural moments of an object, i.e., the way it is composed from parts, and taking them exclusively into consideration; 'object' is here to be understood in the broadest sense. Accordingly, mathematical knowledge can be defined as knowledge based on the structural consideration of objects.[54]

The crucial questions are undoubtedly, what is the extension of *object*, and what kind of objects can be considered or viewed *structurally*. The second question implicitly concerns the boundary between mathematical knowledge secured by intuition (Anschauung), respectively obtained by thinking (Denken) and systematic extrapolation; it is to this question that Bernays turned.

Bernays's analysis of intuitive mathematical knowledge attempts to balance, uneasily, the philosophical demand for intuitive concreteness and the mathematical need for formal abstractness. The tension comes to the fore in first taking formal abstraction as the characteristic feature of intuitive mathematical knowledge, and in then claiming that it is naturally bound to finiteness and finds a principled delimitation only when facing the infinite.[55] It is precisely this coextensiveness of finite and

---

[54] See [7], p. 23. The German text: "Diese Abstraktion, welche als die *formale* oder *mathematische Abstraktion* bezeichnet werden möge, besteht darin, da von einem Gegenstand – 'Gegenstand' hier im weitesten Sinne genommen – die strukturellen Momente, d.h. die Art der Zusammensetzung aus Bestandteilen hervorgekehrt und ausschlielich in Betracht gezogen wird. Man kann demnach als mathematische Erkenntnis eine solche definieren, welche auf der strukturellen Betrachtung von Gegenständen beruht."

[55] The first feature is expressed most clearly on p. 30 of Bernays [7]: "Als das Charakteristische an der mathematischen Erkenntnisweise haben wir die formale Abstraktion, d.h. die Einstellung auf die strukturelle Seite der Gegenstände festgestellt und damit das Feld des Mathematischen in grundsätzlicher Weise abgegrenzt." That is supplemented most forcefully on p. 40: "Die wesentliche Gebundenheit der formalen Abstraktion an das Moment der Endlichkeit macht sich insbesondere dadurch geltend, da bei den Betrachtungen von Gesamtheiten und von Figuren die Eigenschaft der Endlichkeit für den Standpunkt der anschaulichen Evidenz gar kein besonderes beschränkendes Merkmal bildet. Die Beschränkung auf das Endliche wird von diesem Standpunkt

intuitive - and thus the sharp differentiation of the intuitive from the non-intuitive - that was questioned by Bernays himself in the *Nachtrag* to his (1930) [7], when arguing that the epistemological considerations underlying his paper should be revised in light of Gödel's results:

> Of course, the positive remarks, in particular those bringing out the mathematical element in logic and those highlighting elementary arithmetical evidence, are hardly in need of revision. It seems, however, that the sharp differentiation between the intuitive and non-intuitive, as used in treating the problem of the infinite, cannot be carried through this strictly. In this respect then, the view on the formation of mathematical ideas has to be worked out in further detail.[56]

Bernays refers to his later essays in *Abhandlungen zur Philosophie der Mathematik* as containing considerations to address this fundamental issue. The arguments in his (1930) [7] provide, it seems to me, an excellent starting-point, as their detailed examination uncovers revealing difficulties. Thus, I will focus on them without relating them at this occasion to the broader philosophical framework in which they have their systematic place. That framework is deeply influenced by Kant, Fries, and Nelson, with Bernays keeping however a distinctive critical distance. The curious reader may consult Bernays's papers [5], [6], [8]; to recognize that some of the issues discussed below are parallel to (still unresolved) problems in Kant's philosophy of arithmetic, see in particular the writings of Parsons, e.g., [59], [60], [61], and [64].

The arguments that support the uneasy balancing act between philosophically motivated concreteness and mathematically necessitated abstractness and generality are at crucial places strained. That is most evident, when formal abstraction is

---

aus ganz ohne weiteres, sozusagen *stillschweigend* vollzogen. Wir brauchen hier keine besondere Definition der Endlichkeit, denn die Endlichkeit der Objekte versteht sich für die formale Abstraktion ganz von selbst." The second aspect is emphasized on pages 38 and 39, where Bernays argues that formal abstraction helps us to transcend the limits of our "faktischen" or "wirklichen Vorstellungskraft": "An solche Grenzen für die Möglichkeit der Verwirklichung kehrt sich aber die anschauliche [sic!] Abstraktion nicht. Denn diese Grenzen sind vom Standpunkt der formalen Betrachtung zufällig. Die formale Abstraktion findet sozusagen keine frühere Stelle für eine prinzipielle Abgrenzung als bei dem Unterschied des Endlichen und Unendlichen." Bernays continues in the next paragraph: "Dieser Unterschied ist in der Tat ein grundsätzlicher. Wenn wir uns genauer besinnen, wie denn überhaupt eine unendliche Mannigfaltigkeit als solche charakterisiert sein kann, so finden wir, da dieses gar nicht nach der Art einer anschaulichen Aufweisung möglich ist, sondern nur auf dem Wege der Behauptung (bzw. der Annahme oder der Feststellung) einer gesetzlichen Beziehung. Unendliche Mannigfaltigkeiten sind uns demnach nur durch das *Denken* zugänglich. Dieses Denken ist zwar auch eine Art des Vorstellens, aber es wird dadurch nicht die Mannigfaltigkeit als Gegenstand vorgestellt, sondern es werden Bedingungen vorgestellt, denen eine Mannigfaltigkeit genügt (bzw. zu genügen hat)."

[56] See [15], p. 61. The German text: "Freilich, die positiven Ausführungen, insbesondere die Aufweisung des mathematischen Elementes in der Logik und die Herausstellung der elementaren arithmetischen Evidenz, bedürfen wohl kaum der Revision. Jedoch, die scharfe Unterscheidung des Anschaulichen und des Nicht-Anschaulichen, wie sie bei der Behandlung des Problems des Unendlichen angewandt wird, ist anscheinend nicht so strikt durchführbar, und die Betrachtung der mathematischen Ideenbildung bedarf wohl in dieser Hinsicht noch der näheren Ausarbeitung."

supposed to help us in going beyond the limits of our real power of representation (our "faktische Vorstellungskraft" or "tatsächliches Vorstellungsvermögen"); here, *intuitiveness of objects is to be secured by intuitiveness of processes generating them.* A similar step from objects to processes is taken, when Bernays argues next that formal abstraction is essentially bound to the "moment of finiteness". Indeed, Bernays claims, finiteness is not at all a restrictive feature of objects from the standpoint of intuitive evidence: the finiteness of objects is obvious for formal abstraction ("die Endlichkeit der Objekte versteht sich für die formale Abstraktion ganz von selbst"). Why should that be? The answer to this question is given by Bernays paradigmatically for the case of numbers and appeals to their introduction as the "simplest formal objects" by iteration of a successor operation. This intuitive-structural introduction of numbers is appropriate, Bernays claims, only for finite numbers, as repetition is from the standpoint of "intuitive-formal considerations" *eo ipso* finite repetition. In short, and in parallel to the above italicized claim, *finiteness of these objects is to be secured by finiteness of the underlying generative process*.[57]

Finally, according to Bernays, it is the intuitive representation of the finite (die anschauliche Vorstellung des Endlichen) that provides the justification (Erkenntnisgrund) for the principle of complete induction and for the admissibility of recursive definitions, both in their elementary forms. Such a representation of the finite is thus explicitly used, when reflecting on general characteristics of intuitive objects, and it is a crucial presupposition for the proof theoretic approach.

> The *intuitive representation of the finite* [my emphasis, WS] is forced on us, as soon as a formalism is turned into an object of investigation, thus especially in the systematic theory of logical inferences. This brings out that finiteness is an essential moment of any formalism whatsoever.[58]

Bernays's analysis is consequently also basic for other domains whose elements are generated in elementary ways, especially for the domains of syntactic objects needed in proof theoretic investigations. Indeed, he continues by claiming that the limits of formalisms coincide with those of the general representability of intuitive

---

[57] Two aspects are finite: the number of repetitions and, what Bernays calls, the "iteration figure" that formally represents the generating steps of the elementary operation. – Bernays discusses the introduction of natural numbers as the simplest formal objects on pp. 30–2; the argument for the first italicized claim is presented on pp. 38–9, that for the second claim on p. 40; cf. also Note 54. – The importance of this "iterativistic tendency" was emphasized by Hand in his [40] and [41]; cf. also Section 3 of Zach [92]. The suggestion, however, to base a non-standard semantics for numerical statements on this tendency runs into difficulty, when trying to account for the meaningfulness of statements concerning syntactic objects, and that is crucial for proof theoretic investigations. It is an interesting suggestion, if one takes as the "explicit (and only) goal" of the finitist viewpoint "to give an account of truth for (a fragment of) arithmetic which is *secure*", as claimed in [92], p. 44.

[58] See [7], p. 40. The German text: "Zwangsmäig aber stellt sich die *anschauliche Endlichkeitsvorstellung* ein [my emphasis, WS], sobald man einen Formalismus selbst zum Gegenstand der Betrachtung macht, insbesondere also in der systematischen Theorie der logischen Schlüsse. Es kommt hiermit zum Ausdruck, da die Endlichkeit ein wesentliches Moment an den Gebilden eines jeden Formalismus ist."

combinations. (Die Grenzen des Formalismus sind aber keine anderen als die der Vorstellbarkeit überhaupt von anschaulichen Zusammensetzungen.)

How do these considerations compare with Hilbert & Bernays's earlier ones? Should the objects obtained through such elementary generation satisfy the demand articulated in the Notes from 1921/22, namely, that "… the figures we take as objects must be completely surveyable and only discrete determinations are to be considered for them"? - Surveyability was then thought to insure that "our claims and considerations have the same reliability and clarity (Handgreiflichkeit) as in intuitive number theory". Against the backdrop of the generation of numerals we have here the same tension as in the considerations by Bernays, just replacing "surveyability" with "intuitive representability". In order to ground mathematical principles for finitist objects, the elementary and uniform generation of figures has to be appealed to – leading to *purely formal objects* of appropriate abstractness; in order to ground philosophical reflections on the primitive intuitive character of finitist mathematical knowledge, focus is shifted to the surveyability or intuitive representability of individual mathematical objects. Indeed, Bernays claims that we are free to represent (repräsentieren) the purely formal objects by concrete objects (e.g., numbers by numerals) in such a way that these representing concrete objects are intuitable and contain in their structure the essential properties of the represented objects, so that "the relations - to be investigated and holding – between the represented objects are found also between the representatives and can be ascertained by considering the representatives alone".[59]

The deep conflict that is apparent in this intricate discussion is not resolved by argument, but by *fiat*: numbers and other purely formal objects *just are* intuitively given – via representing concrete objects. It is most interesting to observe that Bernays contemplates in his (1934) [9] narrower and admittedly vague boundaries for what is intuitive and distinguishes between numbers that are *reachable* (zugänglich) and those that are not. He does so in a critical discussion of intuitionism and intuitive evidence, viewing as reachable those numbers that do not outstrip our actual power of representation (Vorstellungskraft).[60] He suggests also a way of

---

[59] This is formulated in Bernays [7], fn. 4 on p. 32. The reader should note that "represent" is here actually translating "repräsentieren"; earlier on and later on it is the translation for "vorstellen". The full German text of the note is: "Der Philosoph ist geneigt, dieses Verhältnis der Repräsentation als einen Bedeutungszusammenhang anzusprechen. Man hat aber zu beachten, da gegenüber dem gewöhnlichen Verhältnis von Wort und Bedeutung hier der wesentliche Unterschied besteht, da der repräsentierende Gegenstand in seiner Beschaffenheit die wesentlichen Eigenschaften des repräsentierten Objektes enthält, so da die zu untersuchenden Beziehungen der repräsentierten Objekte sich auch an den Repräsentanten vorfinden und durch die Betrachtung der Repräsentanten selbst festgestellt werden können."

[60] The discussion is found on p. 70. - Parsons makes in his (1982) [60], p. 496, a related distinction for Kant's philosophy of mathematics. He distinguishes between *weak* and *strong* intuitability as follows: "An object is strongly intuitable if it can be intuited, i.e., if it can be an object of intuition. An object is weakly intuitable if it can be represented in intuition without itself being intuitable. This notion is vague because we have not said what is meant by 'representing' an object in intuition. However, representation of abstract objects by concrete objects, or by objects relatively closer

extending mathematical knowledge from reachable numbers to unreachable ones by the *general method of analogy* (die allgemeine Methode der Analogie), i.e., by extending the relations that can be verified for the former numbers to the latter. However one may want to interpret this, it seems clear that finitist mathematics is not secured by intuitive evidence alone. For an adequate conceptual analysis of finitist mathematics one has to go beyond (the representation of) finiteness, admit rather abstract means for capturing the arbitrary finite iteration of elementary steps, and grant in the end for potentially-infinite domains what Bernays asserts for actually-infinite ones, namely, that they can be characterized only by way of a lawful relation (gesetzliche Beziehung). However, the distinctive feature of domains with generated elements is that their lawful relation is not just assumed or claimed, but rooted in our understanding of the underlying generative process; that understanding allows us also to recognize induction principles for proofs and recursion principles for functions.

Bernays's broad *informal considerations* leading up to the natural numbers as *unique* (eindeutige) and *purely formal objects* (distinct from formal objects, i.e. types, that allow different concrete instantiations by tokens) are very appealing and, in a deep sense, similar to those of Dedekind, Helmholtz, and Kronecker.[61] However, only Dedekind who reported on the informal reflections underlying his (1888) [25] in the letter to Keferstein took the further step to a sharp and completely novel mathematical formulation. The latter makes crucial use of infinite sets and in particular of *simply infinite systems*. Following Dedekind, but avoiding infinite sets, Zermelo presented in his (1909) [93] an analysis based on finite sets and "*simply finite systems*". A central question is, *can Zermelo's considerations provide the mathematical basis for a detailed conceptual analysis of natural numbers?* (Bernays's natural numbers as purely formal objects might be obtained then by Tait's "Dedekind abstraction" applied to simply finite systems. Zermelo's work and its connections to that of others is described by Parsons in his (1987) [62].)

5.2 I.d. Classes and Abstract Notions

Hilbert & Bernays's structuralism, when joined with their finitist methodological reflections, is really a structuralism of Parsons's *non-eliminative* variety: it accepts

---

to the concrete, is a pervasive phenomenon and of great importance for understanding abstract objects."

[61] I refer to Kronecker's *Über den Zahlbegriff* and Helmholtz's *Zählen und Messen, erkenntnistheoretisch betrachtet*; both papers were published in the Zeller-Festschrift, Leipzig 1887. Dedekind refers to these two papers when remarking in the first note to the *Vorwort* of his (1888) [25]: "Das Erscheinen dieser Abhandlungen ist die Veranlassung, welche mich bewogen hat, nun auch mit meiner, in mancher Beziehung ähnlichen, aber durch ihre Begründung doch wesentlich verschiedenen Auffassung hervorzutreten, die ich mir seit vielen Jahren und ohne jede Beeinflussung von irgendwelcher Seite gebildet habe."

basic, potentially-infinite domains of constructed objects, in particular, of natural numbers and syntactic figures constituting formalisms.[62] I propose to call their structuralism and extensions thereof *reductive*, because of the special justificatory role the basic structures play. This is in analogy to "reductive proof theory". The systematic connection will become clear, I trust, from the following considerations that concern the challenging question, how to extend the preliminary and not unproblematic reflections of Section 5.1 to appropriate infinitary configurations. Let me make this challenge concrete by describing one paradigmatic result of reductive proof theory. The domains of constructed objects are here the higher constructive number classes.

Brouwer considered in his (1927) [16] infinite proofs[63] and treated them as well-founded trees, i.e., as constructive ordinals of the second number class **O**. The latter are inductively generated according to the following clauses:

0 is in **O**;
if a is in **O**, then the successor of a is in **O**;
if f is a function from **N** to **O** and, for all n in **N**, f(n) is in **O**, then the supremum of the f(n) is also in **O**.

Even higher number classes were inductively defined by Brouwer (and by Hilbert in *Über das Unendliche*); the trees branch over **N** and over previously obtained number classes. These are quite complex i.d. classes, but acceptable to at least some constructivists among them Bishop, Lorenzen, Myhill, but also Martin-Löf. Iterated i.d. classes were at the center of the foundational investigations in the Stanford Report and much subsequent work; see [31]. Church and Kleene [22] had formulated already in the mid-thirties recursive analogues of the higher constructive number classes by requiring that the function f in the third defining clause be (partial) recursive. The elements of **O** can be pictured as infinitary trees that are uniformly and effectively generated; indeed, arbitrary finite subtrees can be effectively determined.

With that specification of constructive function it is quite direct to formulate proof and definition principles for the finite constructive number classes in the language of elementary number theory expanded by predicate symbols for the number classes. The resulting theory, based on intuitionistic logic, is denoted by $\mathrm{ID}^{\mathrm{i}}(\mathbf{O})_{<\omega}$. The paradigmatic result I want to discuss briefly reduces the impredicative subsystem of classical analysis $(\Pi^1_1\text{—CA})_0$ to the theory $\mathrm{ID}^{\mathrm{i}}(\mathbf{O})_{<\omega}$.[64] This reduction is pleasing for two reasons, especially, if one is affected by the implicit irony: (i)

[62] See [64], Section 8.
[63] Brouwer added in a famous footnote: "These mental mathematical proofs that in general contain infinitely many terms must not be confused with their linguistic accompaniments, which are finite and necessarily inadequate, hence do not belong to mathematics." Brouwer claimed that this remark contains his "main argument against the claims of Hilbert's metamathematics".
[64] The result is obtained from work by Tait or by Feferman (in the Buffalo volume (Kino e.a.)[54]) and my 1977 Stanford dissertation [75]; Tait's work goes back to early 1967 and is reported in his (1968) [83]. For the detailed exposition of this and related results see (Buchholz e.a.) and in

$(\Pi^1_1\text{—CA})_0$ suffices as a comprehensive formal framework for the development of mathematical analysis presented in the fourth supplement of Hilbert & Bernays's *Grundlagen der Mathematik II*, and (ii) Brouwer's constructive number classes provide the objective underpinnings for proving the consistency of a blatantly impredicative classical theory. – There is a great deal of contemporary work in proof theory that extends this kind of result mainly by providing ordinal analyses for stronger theories. However, attention has been shifted from subsystems of analysis to a more uniform setting of subsystems of set theory, and the systems of ordinal notations needed for the proof theoretic analysis are connected rather directly with large cardinals in set theory.[65]

Aczel presented in [2] a very general notion of i.d. classes, that is, of classes given by inductive definitions in the broadest sense. All the examples I mentioned (the elementary i.d. classes of terms, formulas, and proofs constituting a formal theory, the Brouwerian constructive ordinals) fall under Aczel's notion. Indeed, Aczel's notion is so general that it encompasses also segments of the cumulative hierarchy of sets. These i.d. classes are in Aczel's terminology *deterministic* and, thus, guarantee the unique generation of the objects falling under them. Given an understanding of the uniform generation steps, the resulting processes allow us to understand the build-up of objects and to recognize proof and definition principles for the domains constituted by them. I call such i.d. classes *accessible domains*[66] and would like to see an abstract mathematical description that highlights their distinctive features. Joyal and Moerdijk's book *Algebraic Set Theory* [53] and the subsequent papers [56], [57] by Moerdijk and Palmgren take, it seems, interesting steps that characterize some classes of accessible domains from the perspective of category theory. Here then is the general question, namely, *can one give a category theoretic characterization of accessible domains?*

It is the crucial task for Hilbert's proof theory to insure the consistency of the "idea of the infinite totality of numbers and of number sets". That is formulated in Bernays [7] in accord with the historical development sketched in Sections 2 through 4, and it required then the use of finitist methods. The task is taken up again in Bernays [9] with a broadened methodological perspective. The assumptions of totalities underlying mathematical theories are called *Platonist*; the condition that restricts their use, as well as the application of the principle of analogy, is described as follows:

> The assumptions we are dealing with amount to representations of totalities and to the principle of analogy or of the permanence of laws. And the condition restricting the application

---

particular my chapter "Inductive definitions, constructive ordinals, and normal derivations" [76], pp. 143–87 of that volume and also [77].

[65] For discussions of this part of advanced proof theory, see [52, 69, 73, 19].

[66] A detailed presentation of the central features of Aczel's i.d. classes is found in Feferman and Sieg [31], pp. 18–25. Elementary i.d. classes are treated in [30].

of these leading ideas is none other than the consistency of the consequences that can be drawn from those basic assumptions.[67]

Accessible domains have a foundational role in providing the means for consistency proofs, whether syntactic, proof-theoretic or semantic, model-theoretic ones. Detailed mathematical and philosophical analyses of accessible domains will allow us to make informative distinctions that concern (constructive) generating operations and their (transfinite) iteration, but also fundamental deductive principles. This leads to a rather natural question, namely, *can theories for accessible domains be given in such a form that their classical versions are uniformly reducible to their intuitionistic variants?* – As I consider suitable segments of the cumulative hierarchy as accessible domains, the investigations of axioms of infinity (i.e., large cardinal assumptions) are of deep conceptual interest and increasingly connected to proof theoretic work, as I mentioned above. This is a part of set theory, where wide-ranging Platonist assumptions in Bernays's sense are being made, and where their consequences on more concrete mathematical problems are being explored. This latter point was emphasized already by Gödel and has been pursued most vigorously by H. Friedman.

Accessible domains reflect the constructive or, if you wish, *quasi-constructive aspect* of mathematical experience; abstract notions like groups, fields, topological spaces reflect its *conceptional aspect*. These two aspects should be contrasted rather sharply. Accessible domains allow us to formulate correct fundamental principles, whereas abstract notions are distilled from mathematical practice to make precise analogies between different areas; that is done for the purpose of comprehending complex connections and obtaining a more profound understanding.[68] I stressed in my (1990) and in [79], pp. 284–5, [78] the broad significance of this distinction; in addition, I argued specifically that the notion of a complete ordered field (characterizing its model, the reals, up to isomorphism) is an abstract one. How different this case is from that of a simply infinite system is revealed by an analysis of the categoricity proofs: in the one case the desired isomorphism follows the build-up of the objects in the domain, whereas in the other the topological completeness has to be appealed to. Thus, there are two complementary important tasks: to analyze the principles for accessible domains and to establish the consistency of abstract notions relative to (theories for) appropriate domains.[69]

---

[67] See [9], p. 75. The German text: "Die Annahmen, um die es sich dabei handelt, laufen hinaus auf Vorstellungen von Gesamtheiten und auf das Prinzip der Analogie oder der Permanenz der Gesetze. Und die Bedingung, welche die Anwendung dieser Leitgedanken einschränkt, ist keine andere als die der Widerspruchsfreiheit der Folgerungen, die sich aus jenen zugrundegelegten Annahmen ergeben."

[68] Bernays makes a related, but certainly not "identical" distinction in his (1930) [7], p. 44, referring in a footnote to Fries; he distinguishes between "dem elementar-mathematischen Standpunkt und einem darüber hinausgehenden systematischen Standpunkt". This more systematic standpoint covers not only analysis, but also set theory, whereas for me set theory with its iterative conception falls under the quasi-constructive aspect of mathematical experience.

[69] In my (1990) I also pointed out that this second task (as well as the first one) cuts across the traditional foundational divides: if there is an abstract notion in intuitionistic mathematics, it is

Let me return to Hilbert. As philosophers, mathematicians, and scientists we should explore Hilbert's broad insights into the complex workings of mathematics instead of keeping him shackled to a narrow foundational position that was taken for programmatic reasons in the twenties. Hilbert's particular proposal for mediating between constructivist and classical positions did not work out. However, the reductive program that emerged from it provides, in my view, an important perspective on mathematical experience. *Reductive structuralism* allows us to connect, in a prima facie coherent way, developments in (the foundations of) mathematics and more directly philosophical studies; it helps us to gain a better understanding of the distinctive character of modern mathematics and its role in our broader intellectual enterprises, in particular its role in the sciences.[70] Hilbert's modernized self can be taken as arguing for creative freedom along two dimensions: *constructions* and *abstract concepts*; the former call for abstract analysis, the latter for constructed models.

6 Concluding Remark

I want to end with a most appropriate comment of Stein's on Hilbert; it mirrors the remark I quoted from Hilbert's Paris Lecture in Section 1. After complaining gently about Hilbert's insistence (in his later foundational investigations) that the statements of ordinary mathematics are meaningless and only finitist statements have meaning, Stein points out:

> Hilbert certainly never abandoned the view that mathematics is an organon for the sciences...; and he surely did not think that physics is meaningless, or its discourse a play with "blind" symbols. His point is, I think, this rather: that the mathematical *logos* has no responsibility to any imposed *standard* of meaning: not to Kantian or Brouwerian "intuition", not to finite or effective decidability, not to anyone's metaphysical standards for "ontology"; its *sole* "formal" or "legal" responsibility is to be consistent (of course, it has also what one might call a "moral" or "aesthetic" responsibility: to be useful, or interesting, or beautiful; but to this it cannot be constrained – poetry is not produced through censorship).[71]

The mathematical (in particular, proof theoretic) and philosophical challenge is, of course, to analyze on what basis we can live up to the responsibility of being consistent.

that of a choice sequence introduced by Brouwer to capture the essence of the continuum; the consistency proof of the theory of choice sequences relative to the theory of the second constructive number class ID(O) can be viewed as fulfilling exactly this task. The proof was given by Kreisel and Troelstra in their paper "Formal systems for some branches of intuitionistic analysis", Annals of Mathematical Logic 1, 1970, 229–387.

[70] Cf. [87], pp. 540ff and [88], pp. 482–4, in particular p. 484 where the earlier ideas are re-expressed.

[71] See [82], p. 255.

Appendix

This appendix contains a note by Bernays that is found in the Hilbert Nachla, Cod. 685:9, 2 and was written, presumably, between 1925 and 1928. It is entitled *Existenz und Widerspruchsfreiheit*. The note is preceded by a brief note on the finitist standpoint (containing only well-known observations) and followed by a note analyzing the criticism of axiomatic set theory by Skolem and von Neumann. For our purposes the latter note is of interest only by the way in which Bernays explicates "consistency of the countable infinite". The consistency proof of arithmetic (including the transfinite axioms for the epsilon operator) establishes the consistency of the countable infinite in the following sense: "An axiom system has been recognized as consistent that cannot be satisfied by a finite system of objects." - Clearly, this topic is taken up in a very illuminating way in Bernays [12].

Existence and Consistency

The claim: "Existence = consistency" can only refer to a system *as a whole. Within* an axiomatic system the axioms decide about the existence of objects.

If, for a system as a whole, consistency is to be synonymous with existence, then the proof of consistency must consist in an exhibition [of a model, WS].

(All consistency proofs up to now have been either direct exhibitions or indirect ones by reduction; in the latter case a certain other system is already taken as existent. – *Frege* has defended with particular emphasis the view that any proof of consistency has to be given by the actual presentation of a system of objects.)

In proof theory, laying a new foundation of arithmetic, consistency proofs are *not* given by exhibition. From this foundational standpoint it does not hold any longer that existence equals consistency. Indeed, it is not the opinion that the possibility of an infinite system is to be proved, rather it is only to be shown that *operating with such a system* does not lead to contradictions in mathematical reasoning [beim Schliessen].

Existenz und Widerspruchsfreiheit

Die Behauptung: "Existenz = Widerspruchsfreiheit" kann sich immer nur auf ein System *als Ganzes* beziehen. *Innerhalb* eines axiomatischen Systems wird über die Existenz von Dingen durch die Axiome entschieden.

Soll für ein System als Ganzes die Widerspruchsfreiheit mit der Existenz gleichbedeutend sein, so muss der Beweis der Widerspruchsfreiheit in einer Aufweisung bestehen.

(Alle bisherigen Wf.-Beweise sind auch entweder direkte Aufweisungen oder indirekte durch Zurückführung, wobei dann ein gewisses anderes System schon als existent angenommen wird. – *Frege* hat bes. nachdrücklich den Standpunkt vertreten, dass der Nachweis der Widerspruchsfreiheit durch wirkliche Aufzeigung eines Systems von Dingen geschehen müsse.)

In der neuen Grundlegung der Arithmetik durch die Beweistheorie geschieht der Wf.-Beweis *nicht* durch eine Aufweisung. Vom Standpunkt dieser Begründung gilt auch nicht mehr, dass Existenz = Widerspruchsfreiheit ist. In der Tat ist ja auch gar nicht die Meinung, dass die Möglichkeit eines unendlichen Systems erwiesen werden soll, vielmehr soll nur gezeigt werden, dass das *Operieren mit einem solchen System* beim Schliessen nicht zu Widersprüchen führt.

Postscriptum

This essay was published originally in *Reading Natural Philosophy*, edited by David B. Malament, Open Court, Chicago and La Salle, 2002, pp. 363–405. That volume of *Essays in the History and Philosophy of Science and Mathematics* was dedicated to Howard Stein on the occasion of his 70th birthday; this re-publication is re-dedicated to him (and appears here with the permission of Open Court).

It has been, and continues to be, my considered view that a penetrating and critical examination of Hilbert's foundational programs and their deep connections to contemporaneous developments in mathematics, physics and logic provides remarkable perspectives for the philosophy of mathematics. I described one such perspective and used it for two purposes: first of all, to articulate a global, integrating approach to foundational work and, secondly, to formulate some more local, focused problems for mathematical investigation. Much fascinating work remains to be done.
Pittsburgh, 12 August 2005

Acknowledgement The translations from the German are mine, unless the references are to English translations; I am grateful that I had access to the Hilbert Nachlass at the University of Göttingen through the work on the Hilbert Edition.

References

1. Ackermann, W., 1924 Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit; Mathematische Annalen 93, 1–36.
2. Aczel, P., 1977 An introduction to inductive definitions; in: *Handbook of Mathematical Logic*, J. Barwise (ed.), Amsterdam, 739–782.
3. Aspray, W. and Kitcher, P. (eds.), 1988 *History and Philosophy of Modern Mathematics*, Minnesota Studies in the Philosophy of Science, vol. XI, Minneapolis.
4. Bernays, P., 1922 Über Hilberts Gedanken zur Grundlegung der Mathematik; Jahresberichte DMV 31, 10–19.
5. Bernays, P., 1928 Über Nelsons Stellungnahme in der Philosophie der Mathematik; Die Naturwissenschaften, 16 (9), 142–145.
6. Bernays, P., 1928a Die Grundbegriffe der reinen Geometrie in ihrem Verhältnis zur Anschauung; Die Naturwissenschaften, 16 (12), 197–203.
7. Bernays, P., 1930 Die Philosophie der Mathematik und die Hilbertsche Beweistheorie; in: (Bernays 1976), 17–61.

8. Bernays, P., 1930a Die Grundgedanken der Fries'schen Philosophie in ihrem Verhältnis zum heutigen Stand der Wissenschaft; Abhandlungen der Fries'schen Schule, Neue Folge, vol. 5 (2), 97–113. (Based on a talk presented on August 10, 1928.)
9. Bernays, P., 1934 Über den Platonismus in der Mathematik; in: (Bernays 1976), 62–78.
10. Bernays, P., 1935 Hilberts Untersuchungen über die Grundlagen der Arithmetik; in: (Hilbert 1935), 196–216.
11. Bernays, P., 1937 Grundsätzliche Betrachtungen zur Erkenntnistheorie; Abhandlungen der Fries'schen Schule, Neue Folge, vol. 6 (3–4), 278–290.
12. Bernays, P., 1950 Mathematische Existenz und Widerspruchsfreiheit; in: (Bernays 1976), 92–106.
13. Bernays, P., 1954 Zur Beurteilung der Situation in der beweistheoretischen Forschung; Revue internationale de philosophie 8, 9–13; Dicussion, 15–21.
14. Bernays, P., 1970 Die schematische Korrespondenz und die idealisierten Stukturen; in: (Bernays 1976), 176–188.
15. Bernays, P., 1976 *Abhandlungen zur Philosophie der Mathematik*; Wissenschaftliche Buchgesellschaft, Darmstadt.
16. Brouwer, L.E.J., 1927 Über Definitionsbereiche von Funktionen; Mathematische Annalen 97, 60–75; translated in: van Heijenoort, 446–463.
17. Brouwer, L.E.J., 1927a Intuitionistische Betrachtungen über den Formalismus; Koninklijke Akademie van wetenschappen te Amsterdam, Proceedings of the section of sciences, 31, 374–379; translated in: (van Heijenoort), 490–492.
18. Brouwer, L.E.J., 1953 Points and spaces; Canadian Journal of Mathematics 6, 1–17.
19. Buchholz, W., 2000 *Relating ordinals to proofs in a more perspicuous way*; to appear in: (size e.a., 2002), 37–59.
20. Buchholz, W., Feferman, S., Pohlers, W., and Sieg, W., 1981 *Iterated inductive definitions and subsystems of analysis: recent prof-theoretical studies*; Lecture Notes in Mathematics, Springer Verlag.
21. Carnap, R., 1930 Die Mathematik als Zweig der Logik; Blätter für Deutsche Philosophie 4, 298–310.
22. Church, A. and Kleene, S., 1936 Formal definitions in the theory of ordinal numbers; Fundamenta Mathematicae 28, 11–21.
23. Dedekind, R., 1872 *Stetigkeit und irrationale Zahlen*; in: (Dedekind 1932), pp. 315–324.
24. Dedekind, R., 1877 *Sur la théorie des nombres entiers algébriques; Bulletin des Sciences mathématiques et astronomiques*, pp. 1–121; partially reprinted in: (Dedekind 1932), pp. 262–96.
25. Dedekind, R., 1888 *Was sind und was sollen die Zahlen*; in: (Dedekind 1932), pp. 335–391.
26. Dedekind, R., 1890 *Letter to Keferstein*; in: (van Heijenoort), pp. 98–103.
27. Dedekind, R., 1932 *Gesammelte mathematische Werke*, Dritter Band; R. Fricke, E. Noether, and Ö. Ore (eds.); Vieweg, Braunschweig.
28. Ewald, W. (ed.), 1996 *From Kant to Hilbert - A source book in the foundations of mathematics*; two volumes; Oxford University Press.
29. Feferman, S., 1981 How we got from there to here; in: (Buchholz e.a.), 1–15.
30. Feferman, S., 1982 Inductively presented systems and the formalization of meta-mathematics; in: *Logic Colloquium '80*, North Holland Publishing Company, 95–128.
31. Feferman, S. and Sieg, W., 1981 Iterated inductive definitions and subsystems of analysis; in: (Buchholz e.a.), 16–77.
32. Fraenkel, A., 1930 Die heutigen Gegensätze in der Grundlegung der Mathematik; Erkenntnis 1, 286–302.
33. Gödel, K., 1933 The present situation in the foundations of mathematics; in: *Collected Works III*, 36–53.
34. Gödel, K., 1938 Vortrag by Zilsel; in: *Collected Works III*, 86–113.
35. Gödel, K., 1986 *Collected Works I*; Oxford University Press, Oxford, New York.
36. Gödel, K., 1990 *Collected Works II*; Oxford University Press, Oxford, New York.
37. Gödel, K., 1995 *Collected Works III*; Oxford University Press, Oxford, New York.

38. Hallett, M., 1994 Hilbert's axiomatic method and the laws of thought; in: *Mathematics and Mind*, A. George (ed.), Oxford University Press, 158–200.
39. Hallett, M.,1995 Hilbert and logic; in: *Québec Studies in the Philosophy of Science I*, M. Marion and R.S. Cohen (eds.) Kluwer, Dordrecht, 135–187.
40. Hand, M., 1989 A number in the exponent of an operation; Synthese 81, 243–65.
41. Hand, M., 1990 Hilbert's iterativistic tendencies; History and Philosophy of Logic 11, 185–92.
42. Hilbert, D., 1900 Über den Zahlbegriff; Jahresbericht der DMV 8, 180–94; reprinted in: *Grundlagen der Geometrie*, 3. Auflage, Leipzig 1909, 256–62.
43. Hilbert, D., 1901 Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900; Archiv der Mathematik und Physik, 3rd series, 1, 44–63, 213–237.
44. Hilbert, D., 1905 Über die Grundlagen der Logik und Arithmetik; in: (Hilbert 1900a), 243–258; translated in: (van Heijenoort), 129–138.
45. Hilbert, D., 1922 Neubegründung der Mathematik; Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 1, 157–177.
46. Hilbert, D., 1923 Die logischen Grundlagen der Mathematik; Mathematische Annalen 88, 151–165.
47. Hilbert, D., 1925 Über das Unendliche; Mathematische Annalen 95, 1926, 161–190; translated in: (van Heijenoort), 367–392.
48. Hilbert, D., 1928 Die Grundlagen der Mathematik; Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 6 (1/2), 65–85.
49. Hilbert, D., 1935 *Gesammelte Abhandlungen*; Vol. 3; Springer, Berlin.
50. Hilbert, D. and Bernays, P., 1934 *Grundlagen der Mathematik*; vol. I, Springer, Verlag.
51. Hilbert, D. and Bernays, P., 1939 *Grundlagen der Mathematik*; vol. II, Springer, Verlag.
52. Jäger, G., 1986 *Theories for admissible sets – a unifying approach to proof theory*; Bibliopolis, Naples.
53. Joyal, A. and Moerdijk, I., 1995 *Algebraic Set Theory*; London Mathematical Society Lecture Notes Series 220; Cambridge University Press.
54. Kino, A., Myhill, J., and Vesley, R.E. (eds.), 1970 *Intuitionism and Proof Theory*; Proceedings of the summer conference at Buffalo, N.Y., 1968.
55. Mancosu, P., 1999 Between, Russell, and Hilbert: Behmann on the foundations of mathematics; Bulletin of Symbolic Logic 5 (3), 303–330.
56. Moerdijk, I. and Palmgren, E., 2000 Well–founded trees in categories; Annals of Pure and Applied Logic 104, 189–218.
57. Moerdijk, I. and Palmgren, E., 2002 Type theories, toposes and constructive set theory: predicative aspects of AST; Annals of Pure and Applied Logic 114, 155–201.
58. Paetzold, H., 1995 *Ernst Cassirer – Von Marburg nach New York*; Wissenschaftliche Buchgesellschaft, Darmstadt.
59. Parsons, C.D., 1980 Mathematical intuition; Proc. Aristotelian Society N.S. 80 (1979–80), 145–168.
60. Parsons, C.D., 1982 Objects and logic; The Monist 65 (4), 491–516.
61. Parsons, C.D., 1984 Arithmetic and the categories; Topoi 3 (2), 109–121.
62. Parsons, C.D., 1987 Developing arithmetic in set theory without infinity: Some historical remarks; History and Philosophy of Logic 8, 201–213.
63. Parsons, C.D., 1990 The structuralist view of mathematical objects; Synthese 84, 303–346.
64. Parsons, C.D., 1994 Intuition and number; in: *Mathematics and Mind*, A. George (ed.), Oxford University Press, 141–157.
65. Peckhaus, V., 1990 *Hilbertprogramm und Kritische Philosophie*; Vandenhoek & Ruprecht, Göttingen.
66. Peckhaus, V., 1994 Hilbert's axiomatic programme and philosophy; in: *The History of Modern Mathematics*, vol. III (Knobloch, E. and Rowe, D.E., eds.), Academic Press, 91–112.

67. Peckhaus, V., 1994a Logic in transition: the logical calculi of Hilbert (1905) and Zermelo (1908); in: *Logic and Philosophy of Science in Uppsala*, Prawitz, D., and Westrestahl, D. (eds.), Kluwer, 311–323.
68. Peckhaus, V., 1995 Hilberts Logik. Von der Axiomatik zur Beweistheorie; Intern. Zs. f. Gesch. u. Ethik der Naturwiss., Techn. U. Med. 3, 65–86.
69. Pohlers, W., 1989 *Proof Theory – an introduction*; Lecture Notes in Mathematics 1407, Springer Verlag.
70. Poincaré, H., 1905 Les mathématiques et la logique; Revue de métaphysique et de morale, 13, 815–35; translated in (Ewald, vol. 2), 1021–38.
71. Poincaré, H., 1906 Les mathématiques et la logique; Revue de métaphysique et de morale, 14, 17–34; translated in (Ewald, vol. 2), 1038–52.
72. Poincaré, H., 1906a Les mathématiques et la logique; Revue de métaphysique et de morale, 14, 294–317; translated in (Ewald, vol. 2), 1052–71.
73. Rathjen, M., 1995 Recent advances in ordinal analysis; Bulletin of Symbolic Logic 1 (4), 468–85.
74. Shapiro, S., 1997 *Philosophy of mathematics – Structure and ontology*; Oxford University Press.
75. Sieg, W., 1977 *Trees in Metamathematics*; Ph.D. Thesis; Stanford.
76. Sieg, W., 1981 Inductive definitions, constructive ordinals, and normal derivations; in: (Buchholz e.a.), 143–187.
77. Sieg, W., 1984 Foundations for analysis and proof theory; Synthese 60 (2), 159–200.
78. Sieg, W., 1990 Relative consistency and accessible domains; Synthese 84, 259–97.
79. Sieg, W., 1997 Aspects of mathematical experience; in: *Philosophy of mathematics today*, E. Agazzi and G. Darvas (eds.), Kluwer Academic Publishers, 195–217.
80. Sieg, W., 1999 Hilbert's programs: 1917–1922; Bulletin of Symbolic Logic 5 (1), 1–44.
81. Sieg, W., Sommer, R. and Talcott, C. (eds), 2002, Reflections on the Foundations of Mathematics, Association for Symbolic Logic, A.K. Peters.
82. Stein, H., 1988 Logos, Logic, Logistiké: Some philosophical remarks on the 19th century transformation of mathematics; in: (Aspray and Kitcher), 238–259.
83. Tait, W.W., 1968 Constructive Reasoning; in: Proc. 3rd Int. Congress of Logic, Methodology, and Philosophy of Science, Amsterdam; 185–199.
84. Tait, W.W., 1981 Finitism; Journal of Philosophy 78, 524–546.
85. Tait, W.W., 2000 Remarks on Finitism; in: (Sieg e.a., 2002), 410–419.
86. van Heijenoort, J. (ed.), 1967 *From Frege to Gödel, a source book in mathematical logic, 1879–1931*; Harvard University Press, Cambridge.
87. Weyl, H., 1925 Die heutige Erkenntnislage in der Mathematik; Symposion 1, pp. 1–32. (Reprinted in vol. 2 of Weyl's "Gesammelte Abhandlungen", Springer Verlag, 1968, 511–542.)
88. Weyl, H., 1928, Diskussionsbemerkungen zu dem zweiten Hilbertschen Vortrag über die Grundlagen der Mathematik; Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 6, pp. 86–88. (Translated in (van Heijenoort), 482–4.)
89. Whitehead, A.N. and Russell, B., 1910 *Principia Mathematica*, vol. 1; Cambridge University Press, Cambridge.
90. Whitehead, A.N. and Russell, B., 1912 *Principia Mathematica*, vol. 2; Cambridge University Press, Cambridge.
91. Whitehead, A.N. and Russell, B., 1913 *Principia Mathematica*, Vol. 3; Cambridge University Press, Cambridge.
92. Zach, R., 1997 Numbers and Functions in Hilbert's Finitism; Taiwanese Journal for Philosophy and History of Science 10, 33–60.
93. Zermelo, E., 1909 Sur les ensembles finis et le principe de l'induction complète; Acta Mathematica 32, 185–193.

Hilbert and the Problem of Clarifying the Infinite

Sören Stenlund

1 Introduction

In Hilbert's presentation of his proof-theoretical program in his paper "On the infinite", from 1926, the program is introduced by connecting its aims to Weierstrass' work on clarifying several problematic notions of the infinitesimal calculus. He says that Weierstrass, "as a result of his penetrating critique, ... has provided a solid foundation for mathematical analysis".[1] What Weierstrass accomplished was to get rid of operations with the infinitely small and the infinitely large by replacing them by operations with the finite without loss of desired results and formal coherence. But what Weierstrass did not do, according to Hilbert, was to clarify the meaning of *the infinite* in mathematics. The infinite is presupposed and comes into play in Weierstrass' work when he is referring, for instance, to *all* real numbers with a certain property. A main task of the proof-theory is said to be to complete Weierstrass' work by giving a clarification of the infinite, and this task is to be accomplished by a strategy similar to the one that Weierstrass used: operations with the infinite must be secured in the finite, otherwise, says Hilbert in a paper from 1931, "the gapless and unified construction of our science would be impossible".[2]

By presenting Weierstrass' work as "a critique", Hilbert indicates that he wants to see Weierstrass' clarificatory work, as well as his own proof theory, as falling within the continuation of Kant's critical philosophy, not in its details but in its general orientation. In a paper from 1922 he says that "just as the philosopher practices the critique of reason; so, in my opinion, the mathematician has to secure his theorems by a critique of his proofs, and for this he needs proof theory."[3] So obviously Kant's critical philosophy, understood as an epistemological project, was a source of inspiration for Hilbert in making mathematical proofs the objects of critical investigation.

S. Stenlund (✉)
Department of Philosophy, Uppsala University, Sweden
Soren.Stenlund@filosofi.uu.se

[1] Hilbert, "On the infinite", p. 183 [16].
[2] Hilbert, "The Grounding of Elementary Number Theory", p. 269 [19].
[3] Hilbert, "The New Grounding of Mathematics", p. 208 [15].

S. Lindström et al. (eds.), *Logicism, Intuitionism, and Formalism,* Synthese Library 341, DOI 10.1007/978-1-4020-8926-8_20,

This Kantian dependence, however, is also one of the problematic features of Hilbert's program. It appears that in Hilbert's thinking certain Kantian ideas of the a priori conditions of knowledge are revised and adapted to early 20th century scientistic views. In any case, a Kantian or neokantian influence can be seen in what Hilbert took to be *the problem of the infinite*. He points out that many confusions and absurdities having their source in the infinite can be found in the literature of mathematics, and he takes the main cause of the problems to be the illegitimate use of certain methods which exceed the limits of contentual (*inhaltliche*) thought. Mathematicians have paid too little attention to the validity of their deductive methods, and to what he calls "the preconditions necessary for contentual logical reasoning."[4] On this point Hilbert refers explicitly to Kant's doctrine that the subject matter of mathematics is given independently of and prior to logic. This is a precondition that he takes to be overlooked in the logicism of Frege, for instance.

The idea of the actual infinite deceives us, according to Hilbert, when we form arbitrary abstract definitions involving infinitely many objects, and it is therefore not a legitimate basis for rational thought. The completed infinite is only a regulative idea of reason in the Kantian sense. We are deceived when we think that in mathematical statements and proofs we can make contentual or material reference to the objects of an infinite totality as real things. The infinite, in the sense of an infinite totality of things existing all at once, is nowhere to be found in reality. He even suggests that the contemporary advances in natural science (in physics and astronomy) gives no support for the idea of the actual infinite, but quite the contrary. We may sometimes think, says Hilbert, that we have encountered the infinite in some real sense, but then we have merely been seduced into thinking so by the fact that we often encounter extremely large and extremely small dimensions in reality. Just as the work of Weierstrass shows that the infinite in the sense of the infinitely large and the infinitely small is merely a figure of speech, so too we must realize, says Hilbert, that "the infinite in the sense of an infinite totality is an illusion."[5]

A similar sceptical attitude to the idea of the actual infinite was expressed by other participants in the debate about the foundations of mathematics in the early twentieth century. That there is no actual infinity was stated explicitly by Poincaré, by Brouwer, by Hermann Weyl, by Thoralf Skolem, and by Oscar Becker. Paul Bernays says that the question of the intuitive knowability of the actual infinite must be answered negatively, and about the same time Wittgenstein says that one of the mistakes he had made in the *Tractatus* was to treat infinity as a number (i.e. as the number of objects of an extension).[6] And Frank Ramsey appears to have converted to a form of finitism in 1929.[7] So there was a striking agreement about the problematic nature of the notion of the actual infinite in mathematics, even though there was

---

[4] Hilbert, "On the infinite", p. 191–192 [16].
[5] Hilbert, "On the infinite", p. 184 [16].
[6] Wittgenstein 1980, p. 119 [36].
[7] Marion 1995b, p. 365 [26].

great disagreement about how the problem should be dealt with and what it meant for the future of mathematics.

What has happened since the 1920s with this problem? My impression is that this problem, as a *philosophical* problem, is not as much alive today as it was in the early twentieth century. But why? Is it mere history and no longer a problem, or at least no longer a serious philosophical issue? Have the problematic issues perhaps been clarified and the problem been resolved?

In the mainstream philosophy of mathematics since the beginning of the 1930s, the problems connected with the actual infinite are not discussed with the same conceptual sensitivity as in the first three decades of the twentieth century. What I have to say in this paper will concern philosophical aspects of this changed attitude as it manifests itself in Hilbert's foundational program and its development after Gödel's incompleteness theorems.

2 Hilbert's Epistemological Claims

There has been much discussion about how Gödel's theorems affect the discussion about the foundations of mathematics that took place in the first three decades of the twentieth century, and in particular how they affect Hilbert's program. I think that convincing arguments have been given for the view that the program cannot be carried out as it was originally intended. Modifications have been developed that are based upon the idea of extending the methods used in meta-mathematical consistency proofs. Other modifications have focused on restricted formal systems in which important parts of classical mathematics can be formalized, and for which consistency can be proved by methods that appear closer to the finitary methods that Hilbert had in mind.

One problem with these modifications is that they do not seriously consider Hilbert's general epistemological aims. They tend to underestimate the extent to which Hilbert saw himself as working within the framework of Kant's general critical philosophy and how central this feature was to his foundational program and formalistic outlook.[8] His distinction between the 'real' and the 'ideal' parts of classical mathematics was inspired by the Kantian distinction between the faculty of the understanding and the faculty of reason, and his idea of resolving the problems of the infinite, including the paradoxes of set theory, was inspired by Kant's resolution of the antinomies of pure reason.[9] There is also a clear similarity between Kant's requirement that the faculty of reason cannot establish any judgement of the under-

[8] See Detlefsen 1993 [7].

[9] Hallett [12], p. 174, remarks that: "We should not, however, make the mistake of thinking either that Hilbert's attitude to the antinomies is simple, or that he underestimated their importance, as is clear from the following passage from a report of his 1905 lecture:

> The paradoxes which we got to know in the above and *which are just a precise mathematical version of the Kantian antinomies* show only too well that an examination of and a new approach to the foundations of mathematics and logic is absolutely necessary." (my italics).

standing that cannot be obtained in principle from the understanding alone[10] and Hilbert's demand that ideal mathematics must be a conservative extension of real mathematics. That Hilbert saw his work on the foundations of mathematics as an investigation of the a priori conditions of the possibility of conceptual knowledge is stated most explicitly in his article *Naturerkennen und Logik*, which was an address he delivered as late as 1930 [18]:[11]

> Philosophers have in fact maintained – and Kant is the classical representative of this standpoint – that besides logic and experience we have a certain *a priori* knowledge of reality. Now I admit that already for the construction of the theoretical framework certain *a priori* insights are necessary and that they always underlie the genesis of our knowledge. I also believe that mathematical knowledge in the end rests on a kind of intuitive insights [anschaulicher Einsicht] of this sort, and even that we need a certain intuitive *a priori* outlook for the construction of number theory. Thus the most general and fundamental idea of the Kantian epistemology retains its significance: namely the philosophical problem of determining that intuitive, *a priori* outlook and thereby of investigating the condition of the possibility of all conceptual knowledge and of every experience. I believe that in essence this has occurred in my investigations into the principles of mathematics.

But as I said before, Hilbert's Kantianism is not unproblematic. There are features of Hilbert's program that are difficult to reconcile with Kant's philosophy.[12] I will not go into the historical and exegetical problems of Hilbert's Kantianism here, but only point out that Hilbert's many references to Kantian philosophical ideas in his writings has been very much ignored in the most influential surveys and commentaries of Hilbert's program. In order to suggest a more fair reading of Hilbert's own writings on the foundations of mathematics, I shall allow myself to take Hilbert at his word when he claims that the epistemological aims, in a general Kantian sense, are a central feature of his work on the foundations of mathematics.[13] In this respect there are significant differences between Hilbert himself and his followers and commentators, and these differences are of great importance for understanding the nature of the changed attitude during the 20th century to the problem of clarifying the infinite.

For someone who is seriously aiming towards a secure philosophical foundation for the science of mathematics, in a Kantian spirit, it makes no good sense to rest content with realizing the aims 'only partly', or in a reductive or relativized sense. Specialized scientific results can be partial or relative, but not so the resolution of

---

[10] Kant, *Critique of Pure Reason*, A328/B385 [20].

[11] Quoted from the English translation in Ewald 1996, pp. 1161–1162 [10].

[12] It is doubtful, for instance, if Kant's critical philosophy can be described as an "epistemology" as this term was used in the late 19th century. One might say with some justification that Hilbert was no Kantian transcendental idealist, but much more a pragmatist or positivist. But then one should not forget that the Kantian critical philosophy was an anti-metaphysical project in which 19th century positivism had its roots. And 19th century positivism is not the neopositivism or logical positivism of the 20th century.

[13] For more details on the relation between Hilbert's epistemology and Kant's philosophy see Kitcher [22] and, especially Detlefsen [7]. The letter article does not seem to have received the attention it deserves among Hilbert scholars.

the philosophical issue of the foundation of scientific knowledge in the a priori conception of philosophy of the Kantian tradition. One of the foremost manifestations of the importance of the philosophical aims for Hilbert is, it seems to me, his repeated insistence on wanting to resolve the problems of the infinite *completely* and *ultimately*. He wanted to bring Weierstrass' critical work on clarifying the infinite in mathematics to completeness, and that could only be done, he thought, by putting classical mathematics on a firm *epistemological* basis by securing the infinite in the finite. "I believe", he said in 1927, "I can attain this goal completely with my proof theory, even though a great deal of work must still be done before it is fully developed."[14] And in a paper from 1931 he went so far as to say: "I believe that in my proof theory I have fully attained what I desired and promised: The world has thereby been rid, once and for all, of the question of the foundations of mathematics as such."[15] Regardless of whether Hilbert was justified in this claim (and I think we must say that he was not), it is clear that the point of the epistemological aspect of the program was to get to the very bottom of the issue, so that it could be resolved *completely*. Anything less than a complete and final solution of the problem would fail to realize Hilbert's original philosophical aims. He saw the confusions about the infinite as the source of the logical paradoxes, and merely to *avoid* existing contradictions was not enough for him: "The chief requirement of the theory of axioms must go farther, namely, to show that within every field of knowledge contradictions based on the underlying axiom-system are *absolutely impossible*."[16] And this requirement is philosophically motivated. In lecture notes from 1910 he says:

> if we set up the axioms of arithmetic, but forego their further reduction and take over uncritically the usual laws of logic, then we have to realize that we have not overcome the difficulties for *a first philosophical-epistemological foundation*; rather, we have just cut them off in this way.[17]

I don't want to say that I take this 'absolutistic' attitude of Hilbert's to be an unproblematic position. On the contrary, I think it is problematic, especially since this goal is to be reached by mathematical means, and I shall return to that issue. But my purpose here is only to point out that the requirement of completeness and finality was a mark of Hilbert's *philosophical* aim, as opposed to specialized scientific ones, a mark of the philosophical orientation in which he wanted to see himself as working, and I think that the seriousness and importance of this attitude in Hilbert's foundational thinking has been underestimated. His aim was not to develop a positive philosophical doctrine about the infinite or anything else, his aim was in a sense the opposite: to liberate mathematics, once and for all, of philosophical worries that he saw as an impediment to the progress of modern mathematics. He wanted to put an end to the restlessness and disputes on foundational issues in the mathematical community, and this restlessness, he thought, would only disappear

[14] See [30], p. 185.
[15] Hilbert, "The Grounding of Elementary Number Theory", in Mancosu 1998, p. 273 [19].
[16] Quoted from the English translation of *Axiomatisches Denken* in Ewald 1996, p. 1112 [14].
[17] Quoted from Sieg [31], pp. 10–11. (my italics).

when the sources of the epistemological worries has been removed completely. As a model for what it would mean to achieve this goal, Hilbert took Weierstrass' clarificatory work in which Weierstrass removed philosophical worries connected with the notions of the infinitely small and the infinitely large.[18]

Weierstrass' work has, however, been taken as a paradigm for a naturalistic conception of philosophical clarification as explication or rational reconstruction. Owing to the influential role this conception of clarification has had in philosophy since the middle of the 20th century, it may be tempting to think that this was also how Hilbert understood Weierstrass' work, and that it was in *that* sense he took it as a model. This is hardly a fair interpretation of Hilbert's philosophical claims. There is no doubt that Hilbert read Weierstrass' work as conceptual clarification in a Kantian spirit.

Kurt Gödel is reported to have commented on the negative consequences of the incompleteness theorems for Hilbert's program as follows:

> What has been proved is only that the *specific epistemological* objective which Hilbert had in mind cannot be obtained. [. . .]
> However, viewing the situation from a purely *mathematical* point of view, consistency proofs on the basis of suitably chosen stronger meta-mathematical presuppositions . . . are just as interesting, and they lead to highly important insights into the proof theoretic structure of mathematics.[19]

Notice that Gödel says that it is '*only*' Hilbert's epistemological objective that is affected, as though this objective was just peripheral in the program, and not at all a central feature. Giving up Hilbert's original epistemological aims, and pursuing the mathematical part of the program is seen as interesting enough!

I think that the attitude expressed in this remark is a common and quite characteristic attitude of the proof theoretic tradition since the 1930s in which elaborations and modifications of Hilbert's program have been developed. The epistemological aims are seen as something secondary that have to adapt themselves to the mathematical research in logic. This is a manifestation of the dominance of mathematical logic in the philosophy of mathematics since the 1930s that constitutes a significant difference to the situation in the first three decades of the twentieth century.

It might perhaps be expected that this development would have led to an increased interest in the philosophy of mathematics among mathematicians, but quite the opposite has taken place. Mainstream philosophy of mathematics since the 1930s has become of less relevance to the concerns of most mathematicians. Kitcher and Aspray [21] point out, correctly I think, that "the distance between the philosophical mainstream and the practice of mathematics seems to grow throughout the twentieth century."[20]

---

[18] Another work, written in a Kantian spirit, that presumably also had influenced Hilbert's view of conceptual clarification was Hertz' *Principles of Mechanics*, where philosophical puzzles about the concept of force are removed.
[19] Quoted from Reid [30], p. 217.
[20] Kitcher and Aspray [21], p. 17.

3 The Epistemological Aim and the Meta-mathematical Program

An idea for modifying Hilbert's foundational aims that has not been discussed very much, as far as I know, is an idea that is rather the opposite of Gödel's: namely, the idea that it is the mathematical part of the original program, *as the means to achieve the philosophical end*, that is the problematic thing; it is the mathematical part that has to be detached from the epistemological objective. The philosophical aim of clarification must be kept alive, but it has to be revised and developed on the basis of the insight that the aim of an ultimate clarification of the nature of the infinite cannot be achieved by means of mathematical construction and proof.[21]

The general philosophical aim must be pursued for reasons that Hilbert expressed in various ways, for instance as follows: "...a science like mathematics must not rely upon faith however strong that faith may be; it has rather the duty to provide complete clarity."[22] What Hilbert suggests here is that nothing can stop our urge to get clarification where there is something unclear and confused, and this holds in particular for the central notions of the science of mathematics which, for Hilbert, is the "the paragon of truth and certitude."[23] We can never accept a situation in which we are forced to give up our aim to get complete and ultimate clarity, and settle with some more or less blind faith in specialized science, in what scientific research may accomplish in an uncertain future.

There is a passage in the article "On the infinite," where Hilbert says:

> ... the definitive clarification of the *nature of the infinite*, instead of pertaining just to the sphere of specialized scientific interests, is needed for *the dignity of the human intellect* itself.
>
> From time immemorial, the infinite has stirred men's *emotions* more than any other question. Hardly any other *idea* has stimulated the mind so fruitfully. Yet, no other *concept* needs *clarification* more than it does.[24]

Hilbert's tone of voice sounds, of course, somewhat old-fashioned and untimely today, but why should we immediately think that the gravity expressed in these words was just some sort of empty rhetorical gesture – as though this lofty way of speaking can only be empty rhetoric? How could someone who expressed himself in this way give up his philosophical objective of trying to get clear about the

[21] This is not a very farfetched idea in the perspective of the Kantian tradition. In the 'transcendental doctrine of method' of the *Critique of Pure Reason*, Kant is very explicit about the methodological difference between mathematics and philosophy, and about the limitations of the mathematical methods in dealing with philosophical issues. The impossibility for any special science (including mathematics) to secure its own epistemological foundation, is a recurring theme also in Husserl's phenomenology, and Hilbert was certainly not unfamiliar with Husserl's ideas, which is not to say that he shared Husserl's views without reservation. Hilbert's student Hermann Weyl was much more sympathetic to Husserl's philosophical outlook and its relevance to the issues about the foundations of mathematics. (See [11]).

[22] Hilbert, "The Grounding of Elementary Number Theory", in Mancosu 1998, p. 268 [19].

[23] Hilbert, "On the infinite", p. 191 [16].

[24] Hilbert, "On the infinite", p. 185 [16].

concept of the infinite, and be content with satisfying merely 'specialized scientific interests'?[25]

But having merged the philosophical and the mathematical aims in the project of a meta-mathematics, the idea of proving the consistency of classical mathematics was after all also a central feature of Hilbert's program. So it might perhaps be objected that giving up the mathematical part would no longer deserve to be called even a modification of the program. It would be a radically different attitude to the problem. The technical-mathematical work in logic that was done during the 1920s by Hilbert and his collaborators and students cannot be detached from Hilbert's program, or from anything that could be called a modification of it. That is certainly true if one adopts Gödel's attitude, as expressed in the quotation above, but the issue depends, again, on how central and important one takes the philosophical aims to have been for Hilbert.

Concerning that issue, it should be noted that terms such as 'finitary', 'finitary means', 'real', 'ideal', 'contentual' (*inhaltliche*) were epistemological, and Hilbert kept them on a philosophical level, and so was his idea about 'the extra-logical concrete objects which are prior to all thinking' as well as the idea of 'securing the infinite on the basis of the finite'. He used them to express the philosophical aims that guides his thinking on the problems of foundations over a long period of time. He speaks of the finitary standpoint as an *attitude* (*Einstellung*), rather than as a technical methodological framework, and of finitary reasoning as reasoning with a special methodological *awareness*. He explains these ideas and notions by means of examples and philosophical elucidations, but he does not give them a final *definition* as technical mathematical notions. They are presented as epistemological notions on a more fundamental level than the mathematical notions and methods that were the objects of investigation. Contemporary commentators often complain that Hilbert is vague and unclear about these notions on the grounds that he does not make them precise by fixing their sense in technical mathematical terms.

I don't want to deny that Hilbert's epistemological notions are in need of clarification, *the issue is rather what clarification should mean here*. Is it self-evident that the technical machinery of contemporary mathematical logic can do the job? Isn't it possible that Hilbert, given his respect for Kant's philosophy, did not quite adhere to the reductively oriented standards of clarification of some of his followers, possibly including his collaborator Paul Bernays?[26] One reason why Hilbert sometimes

---

[25] I agree with Wilfried Sieg that Hilbert was "never too modest about aims" ([31], p. 6), but if we want to understand Hilbert's thinking about the foundations of mathematics I don't think that we can dismiss Hilbert's many statements in his published and unpublished writings on philosophical aims and motives as mere "programmatic" rhetoric. Hilbert was not a philosopher, but a mathematician and scientist who belonged to a generation of pre-World-War European scientists, who still had a deep respect for philosophy as the core of science and European high culture. The opposite attitude, in which science is seen as the core of philosophy, is a more natural attitude among later generations of proof-theorists.

[26] To judge from the appendix to Bernays article "The Philosophy of Mathematics and Hilbert's Proof Theory", ([24], p. 263), an appendix written after Bernays had become acquainted with Gödel's theorems, he seems immediately prepared to revise central philosophical ideas of the

appears vague and unclear may be that he was concerned with larger issues than his followers and commentators were willing to understand.

What cannot be denied, however, is that Hilbert outlined the idea of a proof theory as a kind of contentual mathematical theory, a meta-mathematics, and it is also clear that a basic presupposition or starting-point of this idea was his firm belief that *all existing classical mathematics can be formalized* both in a conceptual sense that was questioned by Poincaré and in a mathematical sense that was refuted by Gödel's theorems. In Hilbert [15] this presupposition is expressed as follows:

> Everything that has hitherto made up mathematics proper is now to be strictly formalized, so that *mathematics proper*, or mathematics in the strict sense, becomes a stock of provable formulae.[27]

In the proof theory classical mathematics exists as a stock of formulas to be made the objects of contentual investigation in the finitary attitude. In a paper from 1928, Hilbert expresses the same idea as follows:

> All the propositions that constitute mathematics are converted into formulas, so that mathematics proper becomes an inventory of formulas . . . The axioms and provable propositions, that is, the formulas resulting from this procedure, are copies of the thoughts constituting customary mathematics as it has developed till now.[28]

What was involved in Hilbert's belief that "customary mathematics as it has developed till now" had been formalized? Panu Raatikainen has argued convincingly that Hilbert must have assumed not only that an ideal mathematical theory is a conservative extension of finitistic mathematics, but also that finitistic mathematics is deductively complete with respect to the real sentences.[29] In an account of Hilbert's ideas, Paul Bernays states it almost explicitly as follows:

> In the case of a *finitistic proposition* [. . .] the determination of its *irrefutability* is equivalent to determination of its *truth*.[30]

If the finitary methods and propositions include most of what can be expressed in primitive recursive arithmetic, then this is refuted already by Gödel's first incompleteness theorem. This is of course a serious challenge to Hilbert's basic idea that all of existing classical mathematics can be formalized as he states it in the quotations just given. How can the formalization of classical mathematics be "copies of the thoughts constituting customary mathematics", if it leaves room for propo-

---

article and accept Gödel's interpretation and to follow Gödel's proposal for modifying Hilbert's program by giving up the original epistemological objective and "extend the finitary point of view." This suggestion is further discussed and developed in Hilbert and Bernays *Grundlagen der Mathematik I and II*, first published in 1934 and 1939, respectively. But both volumes are edited and written by Bernays. I doubt that Hilbert (unlike Bernays) ever could reconcile himself with the notion that proof theory can be at most reductive proof theory

27 See [24], p. 211.

28 Hilbert, "The Foundations of Mathematics", 1967, p. 465 [17].

29 See [29].

30 Bernays, "The Philosophy of Mathematics and Hilbert's Proof Theory", 1998 p. 259 [3].

sitions that are contentual and meaningful in Hilbert's finitary sense, but have no truth-value?

The general idea that existing mathematics had been (or could be) formalized is a fundamental presupposition for Hilbert's notion of proof theory *as a way of realizing his basic philosophical aim*, (that is, the aim of achieving complete and ultimate clarity about the infinite in classical mathematics). In a situation in which the presupposition that existing mathematics can be completely formalized turns out be mistaken, I don't think that it is so very farfetched to imagine that someone (for whom the philosophical aim was as important as it appears to have been for Hilbert) would conclude that *the tools of formal logic and formalization are not adequate for the task of realizing the general philosophical aim*.[31] We cannot get to the bottom of the issue by means of the mathematical tools of proof theory. Proof theory may lead to interesting insights about the structure of mathematics, but as a mathematical enterprise it will only yield clarification in a reductive or relativized sense. To realize the philosophical aim, which is something Hilbert keeps telling us we must, *the problems* of the nature of the infinite must be understood and dealt with in some other way.[32]

---

[31] I think that Poincaré was already convinced of that on conceptual grounds, i.e. on the basis of his understanding of the limitations of formalization (See [28] and [33]). He understood how important the idea of "a first philosophical-epistemological foundation", based upon something like a strict finitism, was for Hilbert's original ideas about "securing the infinite on the basis of the finite". Only then would the infinite in mathematics be secured on the basis of notions and methods that do not already presuppose infinitistic ones. (It appears to me that Hermann Weyl [35] and Oscar Becker [1], who elaborated and defended Poincaré's *vicious circle* argument, also understood the importance of the original epistemological aim for Hilbert.) This ultimate and non-reductive epistemological aim was the main target for Poincaré in his argument that the meta-mathematical project involves a *vicious circle*. It is therefore not surprising that post-Gödelian proof-theorists, who are ignorant of the original epistemological aim, are not impressed by Poincaré's argument. In reductive proof-theory, the issue becomes one of weaker or stronger epistemological significance of proof-theoretical results as measured by weaker or stronger *mathematical* notions and methods used in the proof theory, where the scale of measurement are the categories of mathematical logic. But the original philosophical aims of clarification, if it is to be understood in the perspective of the Kantian tradition, cannot be measured on that scale. It is clarification in a different sense; it is conceptual clarification as opposed to clarification achieved as a result of mathematical construction and proof. What one can question is whether Hilbert himself was enough of a Kantian to understand that difference. His response to Poincaré's argument indicates that he wasn't.

[32] Someone may want to object: "How can it be something we 'must', when it has been proved to be impossible? How can it be our duty to achieve the impossible?" The problem with this objection is that it makes it appear as though there is some lack of clarity about the concept of infinity in mathematics that cannot "in principle" be removed. But it is when the mathematical part of Hilbert's program remains unquestioned *as an instrument for clarification* that it appears as though the goal of complete clarity is impossible to reach. No mathematical result (or scientific discovery) can constitute a limit to the conceptual clarity we can achieve. (The opposite attitude is superstition.) What a mathematical result can do is to change the situation we take ourselves to be in, and thereby also our understanding of our problem. What had been proved shows that one was in a situation different from the one that was taken for granted when the original program was formulated, and, therefore, that the epistemological *problem* was not correctly understood. We can, in a sense, still be in agreement with Hilbert about his philosophical goal of wanting complete

4 What is Classical Mathematics?

One alternative way of approaching the foundational problems was Brouwer's intuitionism [6]. It was an approach that, unlike Gödel's proposal, did not involve bargaining away the general philosophical *aims* of trying to get complete clarification. Why did Hilbert not follow the example of his former student Hermann Weyl, who joined Brouwer already around the beginning of the 1920s? This brings us to the original motive for Hilbert's engagement in the work on the foundations of mathematics: his resistance to the alleged revisionary doctrine of Kronecker and Brouwer, according to which important parts of classical mathematics had to be revised.

But what was the "classical mathematics" that Hilbert was unwilling to revise and wanted to save and secure? At this point I think that we must question the solidity of Hilbert's epistemological aims and realize that they were subordinated to ideological aims concerning the future development of *modern* mathematics. Wilfried Sieg makes a good point when he says:

> The logicophilosophical community has focused on Hilbert's finitist means for securing "classical" mathematics and on the epistemological distinctiveness of those means [...] we have not been equally concerned with the substance of what Hilbert strove to secure – over a lifetime. And that is not *classical* mathematics as it evolved until the nineteenth century, but rather *modern* mathematics as it resulted from a radical transformation during the second half of that century. [...] it was affected mainly by the work of Gauss, Dirichlet, Riemann, and Dedekind.[33]

The main features of the modern mathematics that Hilbert wanted to secure includes the modern axiomatic method, for which the abstract and non-constructive style of Dedekindian mathematics was an important origin. It includes Dirichlet's general concept of function, which was important for the invention of set theory and the logical systems of Frege and Russell. The core of this modern mathematics, developed according to Hilbert's formal axiomatic method, is what we now call "classical formal logic". "Classical mathematics", as the use of this term has established itself in the discussions about the foundations of mathematics since the first decades of the twentieth century, is mathematics, of the past as well as of the present, that is based upon this logic. I shall call this "the foundational sense of classical mathematics".

So, according to the received view within the logico-philosophical community, what Hilbert wanted to secure was classical mathematics (in the foundational sense), and it was classical mathematics *in the same sense* that must be questioned and revised according to Brouwer. Moreover, it is usually taken for granted, within

---

clarity about the notion of the infinite, but then we must admit that the goal cannot be reached along the meta-mathematical paths, and thus that the expression 'achieving complete clarity' must have a meaning different from the one that Hilbert and his followers have taken for granted.

[33] See [32], pp. 364–365.

this community, that most of the practices and results of mathematical work that had been reached at the beginning of the 20th century is classical mathematics, and that most mathematicians working in the established tradition were classical mathematicians.

I think that there is reason to be suspicious of this notion of "classical mathematics". The notion is too conditioned by the positions within the logico-philosophical community of the twentieth century, and in particular by the battle between Hilbert and Brouwer. The notion is conditioned by a questionable presupposition that Hilbert and Brouwer shared and which is built into the notion of "classical mathematics" as it is used in the discussion about the foundations of mathematics since the 1920s. Far from all mathematicians were involved in that discussion, or did take a stand in the issues about the foundations of mathematics. For the mathematicians who did not care about these issues the expression "classical mathematics" may have meant something quite different. The use of the term "classical mathematics" within the discourse of the foundations of mathematics does not signify mathematics of the past seen from an ideologically neutral historical perspective; it involves a *philosophical interpretation* or view of the scientific practices of mathematics, an interpretation the main motivation of which has been to justify mathematics as a deductive science standing on a firm philosophical foundation. The view has its historical roots in Aristotle; its core are the notions and principles of Aristotelian formal logic and philosophy of science, which were transferred to the foundational discourse of the late 19th and early 20th century through the intermediation of Kant.[34] As Michael Detlefsen has pointed out: "In one way or another and to a greater or lesser extent, the main currents of foundational thinking. . .are nearly all attempts to reconcile Kant's foundational ideas with various later developments in mathematics and logic."[35]

Obviously, Hilbert was enough of a Kantian to believe that traditional mathematics, "customary mathematics as it has developed till now", as he called it, *must* be understood according to the philosophical view that is implicit in the notion of "classical mathematics", in which the notions and principles of formal logic have a distinct foundational role (even if not the role that logicists assigned to formal logic). It is this feature of "classical mathematics" that he wants to preserve and extend into modern mathematics, but then in a more formal (and less "genetic") spirit. He tends to express himself in his critique of Brouwer as if this philosophical interpretation of mathematics of the past was somehow a necessary part of traditional mathematics, and as though existing mathematics stands and falls with this traditional philosophical *view* of it.

What we have here is an ambiguity in the use of the term "classical mathematics". In its ordinary, *historical* sense the term should mean something like "mathematical

[34] This core includes not just the classical logical laws, but also the logical notions of a "judgement", "content and truth of a judgement", "the subject-matter of mathematics", "mathematical knowledge as knowledge of the properties of some sort of objects", "mathematical reasoning as deduction", etc.

[35] See [8].

notions, methods and results of the past that are well-established in the practice of mathematics and have a lasting value", but in the discourse of the foundations of mathematics it means roughly: "mathematics as the deductive science that is developed on the basis of classical logic, and in particular on the basis of the law of the excluded middle". In the discourse on the foundations of mathematics, these two senses are assimilated: the foundational sense of "classical mathematics" is taken to be the historical sense.

A weakness of Brouwer's revisionary critique seems to me to be that he essentially shared this view that customary mathematics as it had developed till the beginning of the 20th century was the same thing as what was called "classical mathematics" within the discourse of the foundational discussions.[36] Since classical mathematics, according to that conception, rests essentially on the notions and principles of classical formal logic, a revisionary critique of the use of these notions and principles *had to be* a revisionary critique of customary mathematics as it had developed till that time.

But, again, this use of the term "classical mathematics" does not signify traditional mathematical practice seen from a philosophically neutral perspective. It involves an interpretation, cultivated mainly within the philosophical tradition, and as soon as this is recognized it may very well be possible to have a critique of what is called "classical mathematics" in the foundational discourse that does not involve a revisionary doctrine about *actual* mathematical practice. Indeed, a closer look at certain features of actual mathematical practice, of the past as well as the present, may be a way of showing that the foundational notion of "classical mathematics" involves a philosophical view of mathematics of the past that is in many respects questionable.

I believe that the critique by Kronecker and Poincaré of the use of set theory and formal logic in modern mathematics was a critique of this nature. They did not share the view of mathematics of the past as "classical mathematics" as this term is used in the logico-philosophical community, so, like most mathematicians, Kronecker and Poincaré did not hold positions *within* the foundational discourse as this discourse established itself in the twentieth century. Poincaré was not an intuitionist in Brouwer's sense, but maybe in the sense in which Felix Klein used the term "intuitionist" in his discussion of styles of mathematical thinking.[37] Poincaré

---

[36] This is evident for instance in Brouwer [5], where he speaks of "the imprudent trust in *classical logic*" ([24], p. 49), and presents his version of the interpretation of traditional mathematics as "classical mathematics" based upon classical logic.

[37] I am inclined to agree with what Mathieu Marion ([25], p. 206) suggests when he raises the following question: "Philosophers seldom pay attention to the fact that the early 'constructivists' such as Kronecker or Poincaré professed to have little or no interest in either mathematical logic or the foundations of mathematics. Isn't it possible that in construing Kronecker's (or Poincaré's) constructivism within logical categories, we are missing some insights which were underlying his (or their) so-called 'constructivist' convictions?" As an example one might mention that Poincaré is often concerned with discussing *styles of mathematical thought* in a way that applies to mathematics generally, rather than merely to the foundations of mathematics (for instance in [27]). But due to the dominance of mathematical logic and the foundations of mathematics he is often

attacked features of the different foundational positions from a position outside the foundational discourse, namely on the basis of his experiences *as a reflecting mathematician and scientist*. Many of his general statements in which he expressed his constructive convictions should not be read as philosophical theses or doctrines, but rather as recommendations to mathematicians for a style of mathematical thinking, for how he thinks that good and rigorous mathematics should be done. This is even more true about Kronecker who refrained from using philosophical notions. He rejected, for instance, the use of notions involving completed infinities in algebraic number theory, but not on the basis of some ontological or epistemological position but because he had found them to be unnecessary in his mathematical work.[38]

There is an important difference here between Kronecker and Poincaré on the one hand and Brouwer on the other. In his attacks on Hilbert's program, Brouwer bases his critique on a very idiosyncratic metaphysical-idealist doctrine of the nature of mathematics that is based on the presupposition that mathematics of the past is "classical mathematics" in the foundational sense. So unlike Kronecker and Poincaré, Brouwer developed a position *within* the discourse of the foundations of mathematics – a position that in some respects contained counterparts to the metaphysical view of classical formal logic he opposed. As the logical operations in the classical view are supposed have constant meanings regardless of context, so Brouwer claimed, for instance, to be in possession of a determined concept of existence in mathematics on the basis of which he could judge of any proposed existence proof whether it is correct or not, as though it were decided once and for all what "existence" in mathematics *must* mean. Brouwer's view of existence in mathematics is an external requirement, a dogma, it is not just a recommendation for how to deal with problems of mathematical existence in particular situations.

For Brouwer as for Hilbert, mathematics of the past was "classical mathematics" in the foundational sense. I believe that more unprejudiced historical investigations of the mathematics of the past would show that many, probably most, mathematicians were not classical mathematicians at all in the sense that they took their notions, arguments and ways of reasoning to be guided or justified by formal logic and by the philosophical notions and claims that have surrounded it. For this reason, I

---

erroneously read as though these discussions concern positions in the discourse of the foundations of mathematics. He tends to be read as defending a psychologist position in the foundations of mathematics.

In Klein [23], a distinction is introduced between categories of mathematicians *in general* that he called *logicians, formalists* and *intuitionists* that differs radically from the established classification in the discourse of the foundations of mathematics. The word *logician*, for instance, does not refer to mathematical logic as Klein uses it. Weierstrass is his foremost example of a logician. The issue of styles of mathematical thought appears to be a philosophical issue that has no place in the philosophy of mathematics that is dominated by formal logic.

[38] Hilbert condemned Kronecker as "Verbotsdiktator", but it is not clear to what extent this accusation was fair. There is an interesting revaluation of Kronecker and his work in Edwards 1988. Edwards suggests that "the pendulum is beginning to swing back Kronecker's way, not least because of the appearance of computers on the scene, which has fostered a great upsurge of algorithmic thinking." ([9], p. 142.)

think that the contemporary increasing interest in the history of mathematics will have a great impact on the philosophy of mathematics, provided that it is seriously *historical* investigation and not just accounts of our heritage from "classical mathematics", understood in the foundational sense.

I doubt that most mathematicians' confidence in the correctness of their proofs and results are ultimately a confidence in the general notions and laws of formal logic, even though the outward linguistic form of their arguments might give that impression and may suggest a way of paraphrasing the arguments into the notation of formal logic. Poincaré even suggested that the science of mathematics is a deductive science only in appearance, by which he did not want to deny that high rigour is a distinctive feature of mathematics, but rather that this rigour has to be accounted for in some other way than on the basis of formal logic and formalization. It is the mathematicians' experience and insight into the *specific* subject matter they investigate that constitutes the ground for their confidence in the correctness of their proofs, according to Poincaré. And this insight into a specific mathematical subject matter cannot be captured in formalization and logical inference which is based on notions that are supposed to be common to all mathematical subject matters, and therefore in a sense neutral to any specific one.

5 The Craving for Formal Generality and Its Metaphysical Legitimization

Why has the philosophical interpretation involved in the foundational notion of "classical mathematics" been so widely embraced? What is the source of many mathematicians' belief in it?[39] My hypothesis is that it satisfies *their desire for formal generality and uniformity at the level of the linguistic forms of mathematics*: generality and uniformity across different specific mathematical systems in order to close gaps between them and to create formal coherence and simplicity. Hilbert and his followers have been quite successful in promoting and satisfying this desire. Remember Hilbert's words I quoted at the beginning about the necessity of achieving a "gapless and unified construction of our science." The importance that Hilbert attached to the "method of ideal elements" manifests the same attitude.

But in assigning *foundational* importance to regularities in linguistic forms, certain *general subject matters* have been *invented*, while they are usually presented as *discoveries* in the "classical" view. In order to justify the importance assigned to regularity in forms of expression, contentual or metaphysical correlates to regularity in linguistic forms are needed. This is, in my opinion, the main source of many of the problems connected with the idea of the "actual infinite". Greater awareness and sensitivity towards this tendency, which we can find in some of the foundational

---

[39] Today this belief manifests itself in the distinguished position that set theory is taken to have in mathematics, and in the tendency to use the phrase "foundations of mathematics" as synonymous with "set theory and mathematical logic."

discussions in the first three decades of the twentieth century, particularly in the 1920s, is a more promising starting-point for clarification of the infinite than any meta-mathematical approach.

I am inclined to agree, for instance, with Brouwer [4], in the early critique of the role of language in mathematics – a part of his critique that is quite independent of his positive mentalistic doctrine – that the sources of inspiration for the creation of the problematic infinitistic notions of classical mathematics, and for the extension of the rules of formal logic to new mathematical domains, are similarities in linguistic forms, similarities in the surface grammar of verbal language used in different mathematical systems. Such similarities are taken to signify a constant and common content of a more "abstract and general nature", as it is often expressed. A linguistic form that occurs in the contexts of finite domains (e.g. the expression "for all...") is transferred to, or is recognized to occur also in the context of the infinite, and then the linguistic form is interpreted as having the same, constant meaning in both domains. By taking these, more general, constant meanings to be more fundamental than the meanings of terms belonging to specific mathematical systems, one appears to have closed the gap between different systems, and typically between the finite and the infinite. One of the most obvious examples of this is Russell's idea of taking the general concept of cardinal number as the most basic concept of number, and to define the finite numbers as a special case.

This tendency of "classical mathematics" in the foundational sense, to invent substantial metaphysical correlates as a result of putting too much weight on the sameness, similarities, and regularities in the forms of verbal language used in different areas of mathematics, and paying too little attention to differences in *the forms of their use*, is still quite common in contemporary philosophy of mathematics.

6 The Unity of the Science of Mathematics

There is another, more historical, reason for questioning the foundational notion of classical mathematics and the philosophical view it is built upon. The features of this view, from its original Aristotelian version, that had survived into the 18th and 19th centuries, such as for instance the idea of mathematics as the science of quantity, were not compatible with the extensive change and development of mathematical practice that took place in the 19th century. A new notion of conceptual and formal rigor was needed – one that neither the Aristotelian nor the Kantian philosophical traditions could offer – in order to re-establish mathematics as *pure* mathematics in the form of a unified deductive science standing on a firm epistemological ground. I think that this philosophical legitimization of *modern* mathematics was a main motive for the foundational programs around the beginning of the 20th century. The transitional aspect between the old view with roots in Aristotle and the new view, based on the new lines of development of 19th century mathematics, is obvious in Russell's *The Principles of Mathematics*.

Dedekind, Frege, Russell, Hilbert, Poincaré and Brouwer all express a strong sense of responsibility for the *unity* of mathematics as a science. But it seems to me that they may have belonged to the last generation of great mathematicians in that respect. The enormous development of mathematics into specialties, that in their turn have split up into new specialties during the 20th century, makes it even doubtful if one can say that mathematics today is a science in the traditional sense. Mathematics today is rather a vast field of specialties and research areas that are more or less related to one another, and it is not at all clear how the question about the unity of contemporary mathematics should be answered (except perhaps that all specialties have their "foundation" in the techniques of elementary mathematics). Mathematical logic and the foundations of mathematics have followed the tendencies of their times and gone through this specialization and compartmentalization itself, which may be one important reason why the original foundational motives are no longer alive. In any case, I don't think that the answer to the question of the unity of contemporary mathematics is within the reach of mathematical logic and the foundations of mathematics even though the term "foundations" suggests the opposite. This idea of foundations is rooted in the classical notion of the unity of mathematics as a deductive science that treats of a non-empirical subject matter. I believe that this idea of mathematics as a sort of "natural science of mathematical objects" belongs to the past and that mathematical logic and the foundations of mathematics have found its proper place as a branch of mathematics.

In this situation there is reason for optimism concerning the future of the philosophy of mathematics. Since mathematical logic and the foundations of mathematics have developed into an independent and successful mathematical specialty, there is no longer the need for it to justify itself as philosophy of mathematics and to dominate the philosophical discussion. This will, hopefully, leave the scene open for more sensitivity in conceptual matters, and for neglected philosophical issues of more relevance for actual mathematical practice.

References

1. Becker, O., 1927, Mathematische Existenz, *Jahrbuch für Philosophie und phänomenologische Forschung*, VIII, pp. 439–809.
2. Benacerraf, P. and Putnam, H., (eds.), 1983, *Philosophy of Mathematics, Selected Readings*, Second Edition, Cambridge University Press,Cambridge.
3. Bernays, P., 1930, Die Philosophie der Mathematik und die Hilbertsche Beweistheorie, *Blätter für deutsche Philosophie* 4, pp. 326–367. Translated as "The Philosophy of Mathematics and Hilbert's Proof Theory," in Mancosu 1998.
4. Brouwer, L.E.J., Mathematics and Logic, in Heyting 1975, pp. 72–97 [13].
5. Brouwer, L.E.J., 1928, Intuitionistische Betrachtungen über den Formalismus, *KNAW Proceedings* 31, pp. 374–379. Translated as "Intuitionist Reflections on Formalism" in Mancosu 1998.
6. Detlefsen, M., 1990, Brouwerian Intuitionism, *Mind* 99, pp. 501–534.
7. Detlefsen, M., 1993, The Kantian Character of Hilbert's Formalism, in Czermak, J. (ed.), *Proceedings of the 15th International Wittgenstein-Symposium*, Verlag Hölder-Pichler-Temsky, Vienna, pp. 195–205.

8. Detlefsen, M., 1998, Mathematics, foundations of, in *Routledge Encyclopedia of Philosophy*, Version 1.0, London and New York.
9. Edwards, H., 1988, Kronecker's Place in History, in Kitcher and Aspray 1988, pp. 139–144.
10. Ewald, W., 1996, *From Kant to Hilbert. Readings in the Foundations of Mathematics*, Oxford University Press, Oxford.
11. Feist, R., 2002, Weyl's Appropriation of Husserl's and Poincaré's Thought, *Synthese* 132, pp. 273–301.
12. Hallett, M., 1995, Hilbert and Logic, in M. Marion and R.S. Cohen, (eds.), *Québec Studies in the Philosophy of Science*, Part I, Kluwer, Dordrecht, pp. 135–187.
13. Heyting, A. (ed.), 1975, *Brouwer Collected Works I*, North-Holland, Amsterdam.
14. Hilbert, D., 1918, Axiomatisches Denken, *Mathematische Annalen* 78, pp. 405–415. English translation as "Axiomatic Thought" in Ewald 1996 [10].
15. Hilbert, D., 1922, Neubegründung der Mathematik. Erste Mitteilung, Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität 1, pp. 157–177. English translation as "The New Grounding of Mathematics, First Report," in Mancosu 1998 pp. 198–214.
16. Hilbert, D., 1926, Über das Unendliche, *Mathematische Annalen* 95, pp. 161–190. English translation as "On the Infinite" in Benacerraf, P. and Putnam, H., 1983 [2].
17. Hilbert, D., 1928, Die Grundlagen der Mathematik, *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 6, pp. 65–85. Translated as "The Foundations of Mathematics" in Van Heijenoort 1967.
18. Hilbert, D., 1930, Naturerkennen und Logik, in Hilbert: *Gesammelte Adhandlungen*, Dritter Band, (New York: Chelsea Publishing Company, 1965). English translation as "Logic and the Knowledge of Nature" in Ewald 1996.
19. Hilbert, D., 1931, Die Grundlegung der elementaren Zahlentheorie, *Mathematische Annalen* 104, pp. 485–494. Translated as "The Grounding of Elementary Number Theory," in Mancosu 1998, pp. 266–273.
20. Kant, I., *Critique of Pure Reason*, translated by N.K. Smith, MacMillan, London, 1973.
21. Kitcher, P. and Aspray, W., (eds.), 1988, *History and Philosophy of Modern Mathematics, Minnesota Studies in the Philosophy of Science*, vol. XI.
22. Kitcher, P. 1976, Hilbert's Epistemology, *Philosophy of Science* 43, pp. 99–115.
23. Klein, F., 1911, *The Evanston Colloquium lectures on mathematics*, Macmillan, New York. Partially reprinted in Ewald 1996, pp. 957–971 [10].
24. Mancosu, P., 1998, *From Brouwer to Hilbert, The Debate on the Foundations of Mathematics in the 1920s*, Oxford University[40] Press, Oxford.
25. Marion, M., 1995, Kronecker's 'Safe Haven of Real Mathematics', *Quebec Studies in the Philosophy of Science*, Part I, M. Marion and R.S. Cohen, (eds.), Kluwer, Dordrecht, 1995, pp. 189–215.
26. Marion, M., 1995b, Wittgenstein and Ramsey on Identity, *Essays on the Development of the Foundations of Mathematics*, J. Hintikka (ed.), Kluwer, Dordrecht, 1995, pp. 343–371.
27. Poincaré. A., 1900, Du rôle de l'intuition et de la logique en mathématique. In *Compte rendu du Deuxième congrès international des mathematiciens tenu à Paris du 6 au 12 août 1900*, Gauthier-Villar, Paris, pp. 115–130. English translation as "Intuition and Logic in Mathematics" in Ewald 1996 pp. 1012–1020 [10].
28. Poincaré, A., 1908, *Science et Méthode*, Paris, Flammarion. English translation as "Science and Method", Dover, New York, 1952.
29. Raatikainen, P., 2003, Hilbert's Program Revisited, *Synthese* 137, 157–177.
30. Reid, C., 1970, *Hilbert*, Springer, New York.

[40] Hibert's unpublished lecture notes are all avilable at the University of Göttingen; the reader can find detailed information about them in "David Hilbert's Lectures on the Foundations of Geometry, 1891–1902" (M. Hallett and U. Majer, eds.), Springer Verlag 2004. The information is found in that Volume's section "Hillbert's Lecture Courses 1886–1934", pp. 607-623

31. Sieg, W., 1999, Hilbert's Programs: 1917–1922, *The Bulletin of Symbolic Logic*, Vol. 5.
32. Sieg, W., 2002, Beyond Hilbert's Reach?, in D.B. Malament, *Reading Natural Philosophy*, Open Court, Chicago, pp. 363–405.
33. Stenlund, S., 1996, Poincaré and the Limits of Formal Logic, in J.-L. Greffe, G. Heinzmann, K. Lotenz (eds.), *Henri Poincaré, Science and Philosophy, International Congress, Nancy, France, 1994*, Akademie Verlag, Berlin, Albert Blanchard, Paris, pp. 467–479.
34. Van Heijenoort, J. 1967, (ed.), *From Frege to Gödel: A Source-Book in Mathematical Logic, 1879–1931*, Harvard University Press, Cambridge MA.
35. Weyl, H., 1927, Comments on Hilbert's second lecture on the foundations of mathematics, in Van Heijenoort 1967, pp. 480–484 [34].
36. Wittgenstein, L., 1980, *Wittgenstein's lectures, Cambridge 1930–1932*, from the notes of J. King and D. Lee, Blackwell, Oxford.

Index