Boston Studies in the Philosophy and History of Science 318

Francesca Boccuni
Andrea Sereni *Editors*

Objectivity, Realism, and Proof

FilMat Studies in the Philosophy of Mathematics

Springer

Boston Studies in the Philosophy and History of Science

Volume 318

The series *Boston Studies in the Philosophy and History of Science* was conceived in the broadest framework of interdisciplinary and international concerns. Natural scientists, mathematicians, social scientists and philosophers have contributed to the series, as have historians and sociologists of science, linguists, psychologists, physicians, and literary critics.

The series has been able to include works by authors from many other countries around the world.

The editors believe that the history and philosophy of science should itself be scientific, self-consciously critical, humane as well as rational, sceptical and undogmatic while also receptive to discussion of first principles. One of the aims of Boston Studies, therefore, is to develop collaboration among scientists, historians and philosophers.

Boston Studies in the Philosophy and History of Science looks into and reflects on interactions between epistemological and historical dimensions in an effort to understand the scientific enterprise from every viewpoint.

More information about this series at http://www.springer.com/series/5710

Francesca Boccuni · Andrea Sereni
Editors

Objectivity, Realism, and Proof

FilMat Studies in the Philosophy of Mathematics

Springer

Editors
Francesca Boccuni
Faculty of Philosophy
Vita-Salute San Raffaele University
Milan
Italy

Andrea Sereni
NEtS - IUSS Center for Neurocognition
Epistemology and theoretical Synthax
School of Advanced Studies IUSS Pavia
Pavia
Italy

ISSN 0068-0346 ISSN 2214-7942 (electronic)
Boston Studies in the Philosophy and History of Science
ISBN 978-3-319-31642-0 ISBN 978-3-319-31644-4 (eBook)
DOI 10.1007/978-3-319-31644-4

Library of Congress Control Number: 2016940354

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

To Aldo Antonelli

Preface

The philosophy of mathematics has had a tumultuous life, even when only the last one and half century is considered. While the foundational crisis between the late nineteenth century and the early twentieth century left philosophers with no clear indication of how concerns prompted by reflection on mathematics, its language, and its objects could be eventually solved, mathematics has maintained a central place in philosophical investigations. Led both by empiricist (when not sociologically leaned) approaches attempting at framing mathematics within an overall conception of natural knowledge, and by novel foundational perspectives striving to preserve an autonomous place for mathematical knowledge, the philosophy of mathematics has witnessed a growth of studies since the Seventies of the past century, and is today one of the liveliest and most stimulating areas of philosophical research, where disciplines as diverse as logic, history of mathematics, philosophy of language and science, epistemology and metaphysics find an impressively fertile common ground. While new and provoking positions have been developed on all sides of possible theoretical divides (platonist vs. nominalist, realist vs. anti-realist, empiricist vs. apriorist, philosophy-oriented vs. practice-oriented, and so on), a wealth of investigations has been flourishing, research groups gathering around specific proposals have been forming, and novel perspectives and conceptual tools have been emerging. The essays collected in this volume are meant to be a vivid example of this renewed philosophical milieu, and are the outcome of activities organized by one among the recently born international communities devoted to the philosophy of mathematics.

The background milieu for what would later become the Italian Network for the Philosophy of Mathematics, FilMat, was provided by recurrent meetings of, and collaborations between, national research groups (such as SELP and COGITO), promoted by several researchers in the philosophy of mathematics and logic based at various Universities in Italy (among which the University of Bologna, Scuola Normale Superiore in Pisa, San Raffaele University in Milan, the University of Padua, the University of Milan). Also through the continued encouragement of colleagues like Marco Panza at IHPST (Paris), and a renewed interest in publications in the philosophy of mathematics by Italian publishers (among which we would like to stress the role played by Carocci Editore) participants in those groups came to realize that their network of national and international relations should be given some more stability.

The initial suggestion for the creation of the network was then prompted by a first very successful conference that was organized at Scuola Normale Superiore, Pisa, in 2012, by Gabriele Lolli and Giorgio Venturi. The aim of that conference was to gather Italian scholars in philosophy of mathematics and closely related fields, in order to bring something to the fore: despite the philosophy of mathematics is unduly underrepresented in Italian academia, the community of Italian researchers successfully involved in the discipline is lively and conspicuous. This was not only—and not so much—meant to apply to well-known scholars already based in prestigious universities all around the world, but especially to young scholars, including PhDs and post-docs, who strive to find adequate opportunities in this field in their country, and are most of the time bound to flee towards non-Italian Universities (well beyond what is a legitimate and necessary need for international exchanges and cooperation). The aim of the Pisa conference was to make all Italian researchers, at any level, be they based in Italy or abroad, feel part of a compact and collaborative community, with clear national ties while still extremely well entrenched in international scientific research, with a significant potential for fostering successful careers of young scholars. That potential was displayed in a Springer volume that selected a number of papers from that conference, edited by Gabriele Lolli, Marco Panza and Giorgio Venturi and published in the *Boston Studies for the Philosophy and History of Science* series in 2015: *From Logic to Practice. Italian Studies in the Philosophy of Mathematics.* The volume gave a nice picture of Italian research in the philosophy of mathematics, with a special focus on the integration of logical, historical, and philosophical concerns in the philosophy of mathematics and mathematical practice.

Some of the participants in that conference—Francesca Boccuni, Gabriele Lolli, Marco Panza, Matteo Plebani, Luca San Mauro, Andrea Sereni, and Giorgio Venturi—felt the urge of providing that community with a more solid and visible platform, as a way of acknowledging the reception of Italian researchers in the philosophy of mathematics, of appreciating the successful placement of Italian scholars in international universities, of integrating younger students in a nationally based and yet geographically diffused web of professional connections, and, last but not least, of promoting the undeniable interest of studies in the philosophy of mathematics in Italian academia. Given the disseminated location of the potential members, the network form seemed most appropriate.

The FilMat Network (www.filmatnetwork.com) soon gathered a conspicuous number of affiliations of Italian scholars worldwide, counting almost 70 members at the time this Preface is written. Others may join the network in the future, and we are confident that the number of young students and early-career scholars finding in membership to the network a way of facilitating and enhancing their scientific and professional journey in this field will raise in the long run.

Despite the nation-based nature of the network, nothing is more alien to its promoting ideas and its mission than closure or isolation with respect to an international scientific community which is becoming more and more global and connected. Quite to the contrary, we believe that keeping researchers in this field in close touch despite their diffuse geographical locations is an ideal way of intensifying scientific cooperations independently of national borders, also by building on individual existing collaborations in different countries.

The FilMat Network promotes conferences, workshops, and seminars in the philosophy of mathematics and strictly related areas, also by circulating news about activities organized by its members. One of its main aims so far has been to schedule a biennial network conference. The first official FilMat international conference—*Philosophy of Mathematics: Objectivity, Cognition and Proof*—was organized by the editors of this volume at San Raffaele University, Milan, in May 29–31, 2014. As a way of stressing both the network's nation-based original inspiration and its interest in fostering international cooperation, the conference assumed a specific format. Invited speakers were selected from each of four categories: Italian scholars based in Italy (Mario Piazza, University of Chieti-Pescara), Italian scholars based abroad (G. Aldo Antonelli, UC Davies), non-Italian philosophers (Leon Horsten, University of Bristol), and early-career invited speakers (Francesca Poggiolesi, IHPST Paris). Contributed speakers were selected by double-blind review through an international call for papers.

The conference was—or so we dare say—quite a successful event. It hosted 21 talks—with the addition of an inaugural lecture by Stewart Shapiro (Ohio State University)—and received 38 submissions from 36 international universities and research institutions from Austria, Belgium, Brazil, Bulgaria, Finland, France, Germany, Hungary, Israel, Italy, Poland, Portugal, UK, and USA. Considering joint works, submissions came from a grand total of 42 authors, of which 9 Italy-based and 6 non-Italy-based Italian nationals, and 27 international authors. Among all of them, 18 were young scholars and early-career researchers. Numbers proved the format to be successful, and it is likely to be preserved in the future. The second FilMat conference has taken place at the University of Chieti-Pescara in May 26–28, 2016, organized by Mario Piazza together with Pierluigi Graziani (University of Urbino) and Gabriele Pulcini (University of Campinas), and statistics of submissions in terms of quantity and international provenance equalled those of the first conference.

We are grateful to Springer, to the Editors of the *Boston Studies in the Philosophy and History of Science* series, and in particular to Springer's publishing editor Lucy Fleet for having accepted and supported our proposal of making a volume out of selected contributions from the FilMat 2014 conference, and to the Project Coordinator Karin De Bie, as well as to Steve O'Reilly and Gowtham Chakravarthy for their support in the production process. We also wish to thank two anonymous reviewers for their precious comments on the book proposal and final draft. Together with the collection stemmed from the Pisa conference, we believe this volume will testify the value of the kind of research in the philosophy of

mathematics that has been gathering around activities of the FilMat Network, and we hope other volumes will follow suit.

Above all, we are grateful to all the authors who submitted to this volume for their cooperation and patience in going through a double-blind review process, which has often led to significant modifications and improvements of contributions based on the valuable suggestions from a panel of about 30 international reviewers, to whom we express our warmest thanks. Given changes occurred along this process, and the final distribution of themes across contributions, the title of this book has been slightly changed from the title of the original conference where most papers were originally presented.

We also would like to thankfully acknowledge the support of the PRIN Italian National Project *Realism and Objectivity* (national coordinator: Pasquale Frascolla, Basilicata University), and in particular to the San Raffaele research unit *Cognitive Sciences and Scientific Objectivity* (unit coordinator: Claudia Bianchi, San Raffaele University).[1]

Of all authors who directly or indirectly supported our project—by participating in the conference or submitting their contribution to the volume—this collection is dedicated to G. Aldo Antonelli. Aldo prematurely and unexpectedly passed away on October 11, 2015. He was the nicest person and an outstanding philosopher of mathematics and logic. His death was an immense loss to the scientific community as a whole. Aldo was extremely supportive of the network project since we first invited him as a network member and as a speaker to the FilMat conference in Milan, and maintained his unshaken support for this volume. When he passed away, he was just about to send us a revised version of his paper. Reviewers suggested just minor revisions, and we then decided to publish the paper as it was. In their comments reviewers clearly manifested sincere appreciation for Aldo's paper. Sean Walsh emphasized that "the paper was eloquently composed and a joy to read." Roy Cook stressed that "the paper is excellent," and accompanied his report with the following confidential comment:

> Of course, about halfway through the paper I also became pretty confident of the identity of the author, and if I am right, then my positive report is not surprising: the person who I suspect wrote the paper usually produces excellent work that rarely needs significant modification or revision! (Of course, I could be wrong about who the author is, in which

[1]Even though this volume is a self-standing enterprise, the FilMat conference that made it possible received financial support from various sources. Besides the PRIN project, we take the opportunity to thank again the Ph.D. Program in *Cognitive Neurosciences and Philosophy of Mind* (San Raffaele University/NeTS at IUSS Pavia); the Ph.D. Program in *Philosophy and Sciences of the Mind* (San Raffaele University); SELP (Seminario di Logica Permanente). The conference was held under the auspices of AILA (Italian Association for Logic and its Applications), SIFA (Italian Society for Analytic Philosophy), SILFS (Italian Society for Logic and Philosophy of Science), and in collaboration with COGITO Research Centre (Bologna) and CRESA Research Centre (San Raffaele). Special thanks go to the then Dean of the Faculty of Philosophy at San Raffaele University, prof. Michele Di Francesco, for his support in promoting this and many other activities in the philosophy of mathematics.

case the moral is that the actual author of the paper commendably met the very high standards that are typical for the person who I have in mind).

We are thankful to them for their permission to disclose their names and report parts of their comments as a way of witnessing once more the excellent quality of Aldo's research. We are all the more grateful to Aldo's partner Elaine Landry and his son Federico Antonelli for having made the publication of his paper possible. Aldo was the clearest representative of the kind of scholars the network is thought for, and we are proud to have been given the chance of meeting him and collaborating with him. He was organizing a workshop on *Ontological Commitment in Mathematics* together with Marco Panza, to be held at IHPST in Paris, where he would have presented the paper published in this volume. The workshop was turned into an event *in memoriam of Aldo Antonelli* and took place in Paris on December 14–15, 2015. Andrew Arana, in collaboration with Curtis Franks, delivered a memoir of Aldo's life and work, which they kindly gave us permission to include in this collection. As a way of homaging Aldo's work, together with his partner and his colleagues and friends Robert May and Marco Panza, we decided to include also a discussion note summarizing and systematizing the discussion that took place after Aldo's paper was read at the Paris workshop, which Robert and Marco kindly agreed to edit. We are confident that this discussion will do nothing more than stressing once more how stimulating and thought-provoking Aldo's work in the philosophy of mathematics can be.

If something has to be witnessed by the papers included in this collection, it is the variety of philosophical concerns that may be prompted by current reflection on mathematics. It goes without saying that what is offered here is a necessarily partial picture—as the vast production of papers and books in this field in recent years, supported by the creation of dedicated networks and research groups, testifies. As Stewart Shapiro emphasizes through the consideration of three case studies, there is a variety of stimulating ways in which mathematics and philosophy can reciprocally contribute to an improved understanding of their respective fields. Of these interactions, and more generally of the philosophical concerns that mathematics raises, three are the main areas on which the papers collected here focus, briefly codified in the three key notions in the title: *Objectivity, Realism, and Proof.*

How a pivotal area of our rational life can be granted the objectivity it deserves is a classical problem in the philosophy of mathematics, which becomes extremely pressing when the shadowy nature of its objects is considered and their connection with the concrete, empirically accessible world is investigated. Essays in Part I (*The Ways of Mathematical Objectivity: Semantics and Knowledge*) are all, to different extents, bearing on these issues. Fregean and neo-Fregean philosophy of mathematics attempted to assuage similar concerns by appropriate semantic analysis of mathematical discourse, but shared solutions are a long way off. Aldo Antonelli (as also shown in the Discussion Note of his contribution edited by Robert May & Marco Panza) and Robert Knowles both confront semantic issues concerning mathematical discourse in a Fregean framework. On a different but related note, mathematics can at the same time be thought of as being constituted by *a priori* truths and as both biologically grounded and entrenched in practice and applications. Markus Pantsar and Marina Imocrante investigate in various ways how the alleged *a priori* character of pure mathematics can be integrated with either an empiricist framework informed by cognitive sciences or an epistemology of mathematics especially focused on applications and actual practice.

Objectivity seems assured when mathematics is considered as a discourse about a well-defined realm of objects, which mathematical theories are supposed to describe. However, the nature of different mathematical objects and the structure of the mathematical universe come in a variety of shapes. Essays in Part II (*Realism in a World of Sets: From Classes to the Hyperuniverse*) focus on these issues, with a particular attention to that essential domain of mathematical objects which sets are. Leon Horsten, Brice Halimi, and Gianluigi Oliveri discuss different approaches to the nature of sets, by investigating respectively the import of conceptions of the infinite on a characterization of classes, the relationship between sets and categories, and the conception of set theory as a science of structures rather than individual objects. The remaining essays in this Part, on the other hand, are concerned with finding an adequate picture of the set-theoretic universe, by connecting a realist picture with a pluralist conception of the set-theoretic domains. Claudio Ternullo & Sy-David Friedman, Neil Barton, and Giorgio Venturi all explore, through different approaches, a conception of the set-theoretical universe today known as multiverse, respectively by relating it to the so-called Hyperuniverse program, by investigating the extent to which relativism may be acceptable in a conception of the set-theoretical domain, and by considering how techniques like forcing may support one or another realist view of such domain.

Both the problems of objectivity and realism need to face an undeniable fact: even when it is understood as aiming at a faithful description of an independent realm of mathematical objects, mathematics is a human activity, where the goal of attaining truth is pursued through symbolic languages by regimentation and clarification of more or less informal notions in appropriate formal systems. Essays in Part III (*The Logic Behind Mathematics: Proof, Truth, and Formal Analysis*) offer new perspectives on some classical issues in this vicinity. Contributions by Mario Piazza & Gabriele Pulcini, and Carlo Nicolai, both deal with specific issues concerning truth in formal theories, either as related to our access to the truth of Gödel's sentence $\mathcal{G}$, or as related to the relationship between axiomatic truth theories and comprehension axioms. In the last three essays, the interplay between some central notions in mathematics and metaphysics (including the metaphysics of mathematics) and their proper formalization is explored by Francesca Poggiolesi, Massimiliano Carrara & Enrico Martino & Matteo Plebani, and Samantha Pollock, either by investigating how the proper logic underlying the epistemic and metaphysical notion of grounding should be made precise, or by suggesting that a primitive notion of finiteness may be essential to singling out the standard model of arithmetic, or finally by exploring how informal beliefs may be involved in the appreciation of technical results such as categoricity theorems.

Through their diverse approaches and focus, the essays in this volume collectively prove once more how rich and stimulating mathematics can be for philosophy on its semantic, epistemic, and ontological aspects. They offer novel perspectives on vexed theoretical issues and promise to deepen our understanding of such a fascinating part of human thought like mathematics is. We are confident that they will stimulate further discussion and will greatly contribute to current debates.

Francesca Boccuni
Andrea Sereni

Contents

Objectivity, Realism and Proof in the Philosophy of Mathematics: An Introduction

Francesca Boccuni and Andrea Sereni

Philosophy, Mathematics, and the Philosophy of Mathematics

Since ancient times, mathematics has always been a source of fascination and philosophical reflection. It has traditionally been considered the major example of an area where knowledge can achieve the certainty and exactness many have seen as a human epistemic ideal. Its pervasiveness and usefulness in everyday and scientific applications have made it the prime tool for the study of physical reality. At the same time, both the allegedly heavenly nature of its objects and its dealing with the infinite have become a constant challenge for mundane and finite beings as we are. During centuries, and especially since the nineteenth century thanks to the development of modern logic, mathematics has stopped being just a source of philosophical concern, and has become a powerful instrument in testing philosophical theories on meaning, knowledge, justification (in the form of logical and mathematical proof), and ontology. The philosophy of mathematics has long become a well-defined and still multifaceted area of philosophical investigation, thanks to its many connections with disciplines such as the philosophy of language, logic, epistemology, and metaphysics, not to mention logic and the history and practice of mathematics itself.

Especially during the past century and a half (although in a way that has its roots in views whose development has taken millennia), positions advanced in the philosophy of mathematics have tended to crystallize in a number of oppositions. We

F. Boccuni
Faculty of Philosophy,Vita-Salute San Raffaele University, Milan, Italy
e-mail: boccuni.francesca@unisr.it

A. Sereni
NEtS - IUSS Center for Neurocognition, Epistemology and theoretical Synthax,
School of Advanced Studies IUSS Pavia, Pavia, Italy
e-mail: andrea.sereni@iusspavia.it

find platonists believing in the existence of abstract mathematical entities, opposed by nominalists denying such existence and trying to bring mathematics back to a worldly affair. We find realists believing in the objective mind-independence of the truth-values of mathematical statements, opposed by various sorts of anti-realists or constructivist views, together experimenting variations on the somewhat *cliché* metaphorical dispute between the mathematician as an inventor and the mathematician as an explorer. We find rationalists of sorts trying to secure the *a priori* status of mathematical knowledge, opposed by various brands of empiricists trying to ground mathematics on empirical evidence. We have (or have had, at least) foundationalists looking for the basic bricks of the mathematical edifice, and anti-foundationalists approaching mathematics as a fallible, when not sociologically determined, practice.

In recent times, while some of these oppositions are still standing, much more nuanced approaches have in fact developed. Suffice to think of the sharper tools that empirical findings have offered to long neglected empiricist insights, or even more of the exploration of fruitful connections between traditional philosophical concerns and a more attentive study of mathematical practice (on which also a volume somehow precursor to this FilMat collection is chiefly devoted; cf. Lolli et al. 2015). And yet, many of those antitheses still underlie much of current research. Others can be thought of, also crossing departmental boundaries. One could think, for example, of the contrast between a primacy of philosophical analysis of intuitive notions on the one side, and, on the other, the role of regimentation in formal languages as a means sometimes to explain, other times to replace those intuitive, informal notions. Again, one can think of the opposition between monistic attitudes towards the logic underlying our mathematical reasoning as opposed to more liberal pluralistic approaches. Prompted by significant technical results, this very same rivalry can arise even in different conceptions of the mathematical universe, or universes, itself.

Focusing on three key concepts in the philosophy of mathematics, *objectivity*, *realism,* and *proof*, the essays collected in this volume offer new perspectives on how theories on each sides of those divides can either be defended or opposed, or else on how middle grounds between them can be found. How is the objectivity of mathematics to be secured, even when its subject matter is not given a platonist construal? What semantic analysis should model mathematical meaning when we renounce thinking of mathematical terms as picking out mathematical objects in some univocal and determinate way? Which role should the cognitive roots and the empirical applications of mathematics play in achieving such objectivity? What are the objects, if any, inhabiting the universe of such a fundamental domain of mathematical discourse as that modeled by set theory? And to what extent should we conceive of it as a unique universe, rather than a plurality of universes each regulated by its own axioms? What is the relation between our intuitive conception of truth in mathematics and the limitations imposed by the formal systems possibly required to give rigor and precision to that conception? And more generally, how does formal treatment affect our understanding of many informal notions which seem to regulate mathematical thought?

These are but some of the questions that are raised, and to which insightful answers are offered, in the essays composing this volume. It goes without saying that all the themes that are explored here are nothing but a sample of the amazing richness of the philosophical concerns that are prompted by reflection on mathematics, not to mention the powerful tools that mathematics (and mathematical logic) can offer to philosophical analysis in so many areas.

A vivid picture of this close interaction between mathematics and philosophy is offered by **Stewart Shapiro**'s stimulating opening contribution, *Mathematics in Philosophy, Philosophy in Mathematics: Three Case Studies*. Shapiro first reminds us of the origins of the philosophical fascination of philosophy with mathematics, tracing it back to its roots in ancient Greek thought. By rehearsing Plato's progressive distancing from Socratic method towards the exactitude of thought aided by and modeled on mathematics, we are reminded of how mathematics (including geometry) becomes in Plato's views an essential prerequisite to gain any intellectual advancement, a means to achieve objective knowledge and to draw the soul "from the world of change to reality." Plato's views on mathematics would become essential to theoretical oppositions (rationalists vs. anti-rationalists, platonists vs. anti-platonists, etc.) which are still with us today.

Greek thought is also the culprit of Shapiro's second case study, the birth and growth of logic as a means for understanding thought through what can be seen as a precursor of a regimented formal language in Aristotle's syllogistic. Quickly following the development of logic up to modern times, we are then presented with a nice contrast. On the one hand, the interplay between the exactness of logico-mathematical language and the analysis of ordinary language becomes pivotal in the development of mathematical logic since the end of the nineteenth century, suggesting both that ordinary language could be successfully regimented and controlled, and that mathematics can be the sharpest litmus paper for testing philosophical theories about meaning and knowledge. On the other hand, the possible mismatch between the exactness of logical tools and the possibly loose discourses they are used to systematize became apparent in the analysis of vague terms. Shapiro reminds us of how careful we need be when we use sharp tools to analyze fuzzy subjects, and discusses the effect this warning has had on others' and his own theory of vagueness.

Like vagueness, other paradox-threatening areas of discourse are subject to different attempts at rigorous regimentation by logico-mathematical tools. One is continuity, which underlies paradoxes of motion since antiquity and is central to the development of modern mathematics. As his final case study, Shapiro discusses several conceptions of the continuum, from Greek to modern times, highlighting how they may suggest revisions of our background logic from classical to non-classical ones, and briefly sketches the view Geoffrey Hellman and himself are developing. According to this view, there is no monolithic notion of continuity that can completely capture the intuitive conception we may have of it: there are rather several precisifications, which can be brought to light once the informal notion is regimented through different formal tools.

Shapiro's conclusion is a neat example of how logico-mathematical language can help in the formal analysis of philosophical (and mathematical) notions, and directly connects with the papers in Part III of this volume. More generally, it is easy to see how this opening essay paves the way for the main themes of the contributions to follow: the problem of accounting for mathematical objectivity over and above a commitment to mathematical objects, the problem of giving a satisfying characterization of the mathematical universe, and the problem of adequately capturing informal notions through logical and formal tools. In the rest of this introduction, we will see how all these themes are explored by the authors contributing to this volume, how they deepen our understanding of the relation between philosophy, mathematics, and logic, and suggest novel directions of inquiry for contemporary philosophy of mathematics.

The Ways of Mathematical Objectivity: Semantics and Knowledge

The essays in Part I of this volume all deal, from rather different perspectives, with issues concerned with the objectivity of mathematics (while issues directly pertaining with mathematical realism are postponed to Part II). Despite their difference in focus, they manifest a minimal shared target: they all suggest ways in which it possible to account for the objectivity of mathematics even without endorsing a platonist conception of its subject matter. To this aim, the first two essays focus on the semantics of mathematical discourse, whereas the second two essays explore problems in accounting for knowledge of both pure and applied mathematics. Together, they offer a nice picture of how reflections on meaning and epistemology in contemporary philosophy can offer new perspectives on account of mathematical objectivity.

The Semantics of Mathematical Discourse

It is has become a somewhat abused habit to begin discussions of the relationship between realism and objectivity by quoting the old dictum attributed to Georg Kreisel according to which the central concern in the philosophy of mathematics is not the existence of mathematical objects, but rather the objectivity of mathematical truths. If we indulge in this habit it is not (just) for lack of imagination, but mostly because the first essay in this section is a compelling contemporary example of how that dictum can be given new life.

It is a well-known component of Frege's philosophy of mathematics that numbers are to be conceived as abstract objects, more specifically a particular kind of logical objects, i.e., extensions of concepts. Since (at least) Benacerraf (1973), platonism has been on the receiving end of a powerful criticism: it is hard to reconcile its answer to the question "what are mathematical objects?" with a plausible answer to the question "how can we have knowledge of them?"

Neo-logicists like Crispin Wright, Bob Hale and collaborators—who preserve the platonist component of Frege's philosophy, freed from any appeal to extensions (cf. Hale and Wright 2001)—have resorted to a suitable epistemology of abstraction principles in order to account for how knowledge of natural (finite cardinal) numbers is achieved. Still, some view the appeal to abstract entities, even when made in the context of consistent principles such as Hume's Principle, as unwelcome. **Aldo Antonelli** also takes it as unnecessary. Building on previous works (see e.g., Antonelli 2010a, b, 2013), in his paper *Semantic Nominalism: How I Learned to Stop Worrying and Love Universals* he offers a naturalistic view of abstraction principles which is claimed to be epistemically more adequate than the neo-logicists' one. In his view, abstraction principles do not provide us with a particular system of objects (cardinal numbers), but rather with representatives for equivalence classes of second-order entities. Such representatives will be available provided the first- and second-order domains are in the equilibrium dictated by the abstraction principles, but otherwise the choice of representatives is unconstrained. Under this conception of abstraction, abstract entities are the referents of abstraction terms: such a referent is to an extent indeterminate, but we can still work with such terms, quantify over their referents, predicate identity or non-identity, etc. Our knowledge of them is limited, but still substantial. In particular, we know whatever has to be true no matter how the representatives are chosen, i.e., what is true in all models of the corresponding abstraction principles. We won't know anything about the special nature of the representatives. But we will know whatever follows from the positing of such representatives. Two remarkable features of this proposal should be noticed. First, it is backed up by an "austere" conception of universals, according to which these are first-order objects, i.e., ways of collecting first-order objects. In this, Antonelli builds on Dedekindian suggestions in order to claim that second-order logic does not by itself import any novel ontological commitment over and above an ontology of first-order, naturalistically acceptable, objects. Second, it is a striking outcome of Antonelli's proposal that, in the case of arithmetic, even if Hume's Principle is understood under the construal described above, the proof of Frege's Theorem (the derivation of second-order Peano Axioms from full impredicative second-order logic with the sole addition of Hume's Principle) still goes through unaffected, since there is nothing in the proof that depends on an account of the "true nature" of numbers. Thus, we are left with a viable construal of logicism, given by the combination of semantic nominalism and a naturalistic conception of abstraction. It goes without saying that this proposal is thought-provoking and open to developments and objections, as the lively discussion of Antonelli's paper held by a number of scholars at a recent Paris workshop on *Ontological Commitment in Mathematics* (IHPST, December 14–15, 2015) and reported in the *Discussion Note* edited by **Marco Panza** and **Robert C. May**, clearly displays.

Antonelli's view is moved, at least partly, by concerns with an ontology of abstract mathematical objects which seems to be suggested by a certain reading of abstraction principles. The same platonist ontology seems suggested by a rather natural semantic analysis of ordinary attributions of numbers, i.e., those statements like “The number of planets in the solar system is eight” (*Zhalangaben*, in Frege’s original language). Following Frege, this analysis interprets such statements as identity statements of the form “The number of planets in the solar system = 8.” The same holds for statements attributing measures like “The mass of Jupiter in kilograms is 1.896 x 10^{27}.” In his paper *Semantic Assumptions in the Philosophy of Mathematics*, **Robert Knowles** discusses and rejects this analysis. Rival accounts of this kind of sentences have been offered (e.g., by Hofweber 2005 and Moltmann 2013). Knowles considers a wide range of linguistic evidence and finds these alternative approaches defective. Any suitable analysis must account for the fact that numerals (and expressions for magnitudes) can occur in natural language both in substantival and adjectival position. Only the former can lend support to a realist conception of mathematical objects, since the semantic function of adjectival expressions is not to pick out singular objects. Frege famously held that sentences in which such expressions occur in adjectival position can always be transformed in sentences in which they occur in substantival position, but many authors have found this claim questionable. Knowles agrees with views rival to Frege’s that the pre-copular expressions in sentences such as the above are not referring expressions. Building on an account of interrogatives, he suggests that a sentence like ‘The number of planets in the solar system is eight’ is true if and only if the fact that uniquely and exhaustively answers the question ‘How many planets are there in the solar system?’/‘What is the number of planets in the solar system?’ is identical to the fact that there are eight planets in the solar system. Likewise for attributions of measures. Thus specified, the truth conditions of these sentences do not seem to involve mathematical objects, but only certain kinds of facts. By itself, then, their truth cannot underlie any realist argument concerning mathematics objects, at least not until either evidence from the semantics of other mathematical sentences is offered, or additional arguments are presented to the effect that it is in the very nature of the facts making those sentences true that mathematical objects are among their constituents.

Mathematical Knowledge, Pure and Applied

Knowles’ discussion suggests that, at least in so far as we focus on statements of applied mathematics, there may be (linguistically adequate) ways of accounting for their objective truth-value which do not resort to any appeal to the role of abstract mathematical objects. The problem of salvaging the objectivity of mathematics while renouncing the natural and yet problematic picture of an alleged acquaintance with mind-independent shadowy abstract entities has to account for at least two salient aspects of mathematics. First, mathematics is a human activity deeply entrenched in our basic cognitive capabilities; and second, one major reason to believe in the truth of mathematics comes from its impressive record of successful applications. The next two essays in this section focus on these two aspects of mathematics respectively.

In *The Modal Status of Contextually A Priori Arithmetical Truths*, **Markus Pantsar** discusses how some crucial epistemic features of arithmetical knowledge, such as aprioricity, objectivity, and necessity can be accounted for in a view of mathematics that gives justice to its biological roots. An "empirically feasible" philosophy of arithmetic (cf. Pantsar 2014), according to Pantsar, sees it as stemming from a bundle of biological primitives regulating basic numerical skills (as recently detailed by findings in the cognitive sciences), a "proto-arithmetic" which later develops into actual mature arithmetic thanks to the role of language and the understanding of a successor operation. This empirical conception of arithmetic can dispense with an appeal to abstract objects, and still claim for an *a priori* character of arithmetical knowledge. Arithmetical knowledge is explained to be "contextually *a priori*": once empirical facts determine a particular context, arithmetic is *a priori* in so far as its methodology is detached from those of empirical sciences and its subject matter is not given by the psychological processes of mathematical reasoning. Biological primitives underlying mathematical knowledge also constrain our ways of experiencing the world in a way that seems to afford the required objectivity (or "maximal intersubjectivity") of arithmetic. Pantsar then focuses on arguing that in his framework arithmetical truths come out as necessary too. Since such truths are grounded in biological primitives of contingent beings, it seems they cannot be true in all possible words, and thus numerical terms cannot function as rigid designators. However, they can rigidly designate those concepts which are developed through our biological settings in all those possible worlds which are inhabited by sufficiently developed biological beings, thus picking out the same thing in all those worlds in which that thing exists. And this, argues Pantsar, is enough to bestow on arithmetical truths their necessary modal status.

While Pantsar deals with reconciling traditional epistemic features of arithmetic with a cognitively informed account of its development, **Marina Imocrante** focuses on how accounts of applied mathematics can foster our understanding of mathematical knowledge. In her *Epistemology, Ontology and Application in Pincock's Account* she critically discusses one of the most developed proposals for an epistemology of applied mathematics, the structural account advanced by Pincock (2012), according to which, roughly, applications of mathematics can be explained through structural relations (mappings or morphisms) between physical systems and suitable mathematical structures. In Pincock's account, applications can be explained with no prior commitment to the ontology of pure mathematics, which has to be decided independently. However, according to Imocrante, Pincock makes a number of assumptions that together threaten to make his account unstable. On the one hand, he offers an "extension-based epistemology" for mathematical concepts, which is meant to take into proper account the historical development of such concepts in mathematical practice. On the other hand, he recommends that pure mathematical statements should be justified *a priori*. Pincock couples these claims with the adoption of a form of semantic realism for mathematical statements and a form of semantic internalism for mathematical concepts. Imocrante suggests that the latter seems to stand in contrast with the aforementioned extension-based epistemology, and should rather be replaced by a form of externalism about mathematical concepts. As a consequence, the structural account would be dispensed by the need of requiring a *a priori* justification for mathematical statements. In Imocrante's account, however, semantic externalism for mathematical concepts does not entail commitment to a form of ontological realism about mathematical objects. On the contrary, Imocrante suggests possible ways in which such externalism can be made consistent with a "world-driven" understanding of mathematical concepts as determined by contingent facts in the history of mathematical development, consistently with the extension-based epistemology underlying the structural account of applied mathematics.

Again, mathematical discourse seems to gain the required objectivity in ways that do not require thinking of mathematics as a true description of a realm of *sui generis* abstract mathematical objects. All essays in Part I offer different perspectives on how to reach this aim, through both semantic and epistemological analysis of pure and applied mathematics. However successful, all these views have to cope with the problematic but still very natural intuition that it is such a domain of immaterial objects that mathematical theories may be about. While many believe that there are various routes to avoid similar commitments in the case of the theories of natural and other numbers, it is much harder to demise the realist picture when we come to more fundamental theories such as set theory. The essays in Part II are devoted to explore different aspects and diverging conceptions of how the universe of sets may nonetheless be characterized.

Realism in a World of Sets: From Classes to the Hyperuniverse

Even when the picture of mathematics as a discourse aimed at a true description of a self-subsistent, mind-independent domain of abstract objects is endorsed, it is far from clear how such a domain should be thought of. Several reductive strategies may be available (be they successful or not) when it comes to objects like natural or real (or complex, for what matters) numbers. In an Ockhamist spirit, some may want to impoverish the apparent abundance of the mathematical universe to a single kind of fundamental objects, to which others can be suitably reduced. Set theory has historically been playing the role of such mathematically and ontologically fundamental theory. Still, a proper understanding of both these reductive strategies, and of such fundamental domain of mathematical objects, requires us to have a clear picture of how the set-theoretic universe is structured. Several problems opens up here, different sets of set-theoretical axioms can be explored and proposed through different strategies, relations between sets and other mathematical objects like classes, categories and structures may be assessed, diverging conceptions of the set-theoretic universe may be forthcoming, and it may even turn out that realism about such a unique universe delivers a too simplistic picture, which should give way to some sort of pluralist conception. The first three essays in Part II focus on crucial problems concerning the notions of class and absolute infinity, the relation between sets and categories, and a platonist versus a structuralist conception of the subject matter of set theory. The remaining three essays deal with pictures of the set-theoretic universe alternative to the monist realist one, in conceptions known as multiverse and hyperuniverse, and with how new axioms should be justified in developing such a plurality of universes.

Varieties of Mathematical Objects: Classes, Categories, and Structures

Are there collections beyond sets? As **Leon Horsten** reminds us in his *Absolute Infinity in Class Theory and in Theology*, Zermelo thought that there are none, that the set-theoretic universe is made of a potentially infinite hierarchy of so-called normal domains, and that the set-theoretic universe itself cannot be a completed collection, cannot be considered as a set and cannot be quantified over. Opposite to Zermelo's view seems to stand Cantor's conception of the set-theoretic universe as a completed absolute infinity. Horsten notices how Cantor's conception of absolute infinity recalls some conceptions of the infinite in Western theology. The role of reflections principles underlies this analogy: once the whole set-theoretic universe can be reflected and represented by some collection at a lower stage, it becomes somehow ineffable, and ineffability is one of the traits of absolute infinity in theology. Indeed, Cantor himself is not clear on whether his conception of the absolutely infinite is theological or mathematical in nature. Horsten explores the tensions in Cantor's conception, and possible different interpretations of his view, but aims at rehabilitating it in contrast with Zermelo's. If we adopt Cantor's view that the universe is a completed whole and acknowledge classes beyond sets, Horsten shows, we can motivate stronger reflections principles than what is allowed in Zermelo's framework, like the Global Reflection Principle (*GRP*). This very roughly states that the whole universe with its parts is indistinguishable from some initial set-sized cut of itself and its parts. And this idea closely resembles Philo of Alexandria's view that "there are angels such that every humanly describable property of God also applies to them." After advocating a mereological conception of proper classes, Horsten distinguishes between a mathematical global reflection principle, $GRP_{\sum^0_\infty}$, where only quantification over sets is allowed, and a mereological one, $GRP_{\sum^1_\infty}$, where also quantification over proper classes is allowed. In the resulting picture, "good non-theological sense" is made of Cantor's picture of the set-theoretic universe. This view not only allows for stronger reflection principles and large cardinal axioms, but also allows claiming that such principles are "intrinsically motivated" by the same pattern of reasoning that justifies analogous principles concerning the absolutely infinite in theology.

Horsten's essay presents us with competing conceptions of how the cumulative hierarchy of sets could be conceived and the axioms describing it justified. As we will see in the rest of this part of the volume, several pictures of the set-theoretic universe can be contrasted. But one may also wonder whether sets are indeed the right entities to play the crucial foundational role they have been long thought to play. Traditionally, that role has been challenged by category theory (cf. MacLane 1978). As **Brice Halimi** initially reviews in his *Sets and Descent*, a host of arguments have been offered to decide the rivalry between sets and categories (from claiming that categories, as collections of objects plus collections of arrows, presuppose set theory, to claiming that *ZFC* is nothing but by a particular case of set theory that can be developed using an elementary theory of the category of sets). However, Halimi's intent is not to have a final say on the dispute, but rather to assume a mathematically informed attitude through close looks at mathematical practice, and to investigate when set theory and category theory can combine in fruitful ways, well beyond the adjudication of the foundational primacy to either. Halimi focuses on one relevant instance of this possible interaction (whose technical details are perspicuously explored in the course of the exposition): Algebraic Set Theory (AST), a reconstruction of *ZFC* in category-theoretic terms inspired by algebraic geometry (cf. Joyal and Moerdijk 1995). More specifically, Halimi shows how AST is guided by descent theory, a theory coming from algebraic geometry which studies the shift from local data to a global item through a "glueing" procedure. Halimi first introduces the framework of fibered categories (a category-theoretical generalization of the notion of surjective maps), a notion on which descent theories relies. He then proceeds by showing how the first axioms of AST combine the respective frameworks of both *ZFC* and descent theory, and concludes by stressing how AST neatly displays a fruitful interaction between the two theories. In a nutshell, AST uses a fibered category in order to reinterpret *ZFC* as an arrow-based theory and to enrich it with the geometric ideas of localization and glueing. So the way in which AST exploits both category and set theory is grounded in techniques coming from abstract algebraic geometry and algebraic topology. Beyond exemplifying a nice cooperation between the two rival theories in the foundation of mathematics, then, AST also has the advantage of linking those foundational theories to other crucial portions of mathematical practice.

Whether or not one follows Halimi's advice of focusing more on the interaction than on the foundational rivalry between sets and categories, it is undeniable that finding a proper characterization of the set-theoretic universe has had a pivotal function in essentially foundational projects. The pull towards the intuitive conception of that universe as a unique totality (which may not itself form a definite collection) is admittedly strong, and reinforced by views of that universe (as the one due to Cantor) that we have already seen explored by Horsten. According to what **Gianluigi Olivieri**'s discussion in *True V or not True V, That is the Question*, this static picture of the universe is wrong: there is no such thing as the one true universe of sets. According to Oliveri, this naïve idea collapses as soon as one investigates further the accompanying thought that set theory is a science of objects. This generates a two-horn dilemma. On the one hand, if first-order *ZFC* is consistent, and *V* is a model of it, then we are left with the consequence (by Gödel's theorems, Löwenheim–Skolem theorem, and forcing techniques) that there is a plurality of even non-isomorphic such models, and we are at a loss in individuating the one true universe. On the other hand, if we renounce the idea of a one true universe *V*, we are, according to Oliveri, bound to adopt views (such as constructivism or Meinongianism about set-theoretic objects) that negatively affect both our treatment of independent questions like *CH* and the foundational role of set theory more generally. Oliveri's way out of the dilemma consists in rejecting the idea that mathematics is a science of objects in favor of the view that mathematics is a science of structures. This latter view finds unacceptable the very idea of a universe as the totality of all sets and, therefore, is against the idea of a true *V*. It also upholds metaphysical realism about structures, though it does not do away with mathematical objects, but merely restricts mathematical investigation to the study of structures. Contrary to what happens with other views renouncing the idea of a one true universe, however, it is committed to realism about truth-values of mathematical statements. This picture of the subject matter of set theory, in the end, is intended to oppose the "architectonic metaphor" which, according to Oliveri, underlies so much of the discussion of set theory as a foundation of the edifice of mathematics, offering a much more nuanced view of foundational issues in mathematics, coherently with a view of the latter as a partly fallible and conjectural discipline, which Oliveri motivates and supports.

Varieties of Mathematical Universes: Multiverse and the Hyperuniverse
The first three papers in Part II raise specific questions concerning the universe of sets conceived as the single domain of a foundational theory: whether it should contain only sets or also classes when accounting for absolute infinity, whether it should be conceived algebraically leading to new perspectives on the relations between sets and categories, and whether such domain is properly characterized as being a domain of objects and not of structures instead. The last among these essays suggested that the question of what is the one true set-theoretic universe may be an ill-posed question. Without abandoning a conception of set theory as a science of objects, the authors of the remaining three papers in Part II may share the same concern. As a consequence, they explore different views of the nature of the set-theoretic universe, or universes. Multiversism and the Hyperuniverse picture building on it suggest that different systems of objects obtained by variously interpreting the axioms of set theory could be taken to constitute a plurality of distinct and still coexisting universes. When further investigated, this conception elicits a vast array of concerns for any realist attitude towards sets, and for our understanding of axioms as basic descriptions of a univocal domain of objects.

In their contribution *The Search for New Axioms in the Hyperuniverse Programme*, **Sy-David Friedman** and **Claudio Ternullo** explore a novel procedure for the search of new intrinsically justified axioms in the Hyperuniverse program recently developed by Arrigoni and Friedman (2013). The authors first distinguish between potentialist and actualist conceptions of the set-theoretic universe, and review Zermelo's conception as being potentialist in height (the height of *V* is not fixed and new ordinals can always been added) and actualist in width (the width of *V* is fixed and no new subsets can be added at each stage). This conception gives rise to a form of vertical multiverse that only partially (in its height dimension) satisfies a principle of plenitude which the authors take as underlying the iterative conception of set, according to which "given a universe of sets, all possible extensions of it which can be formed are actually formed." The Hyperuniverse ($\mathbb{H}^{ZFC}$) is a conception of the set-theoretic universe which is meant to allow for maximal extendibility in both dimensions: it is the collection of all countable transitive models of *ZFC*. Friedman and Ternullo motivate the restriction to such models and explore the underlying logic in which satisfaction in this multiverse can be defined. They then move on to assess new candidate axioms, which now take the form of higher-order maximality principles about V, formulated in a Zermelian framework, satisfied by members of $\mathbb{H}$, and coming with a stock of associated first-order consequences. Rival conceptions of what new set-theoretical axioms should be are then discussed. As far as the background ontology is concerned, Friedman and Ternullo call their view "dualistic," meaning that they endorse elements of both monism and pluralism about the set-theoretic universe: they postulate one single, maximally extendible universe, but they also countenance different universes given by the relevant models, where new set-theoretical truths can be detected.

Friedman and Ternullo are chiefly moved by epistemological concerns. They want to suggest novel evidential paths to secure truth of new axioms without endorsing any pre-formed ontological picture. Accordingly, their view is more focused on how the concept of set should be adequately cashed out, rather than with how a particular view of set-theoretic ontology should be motivated. It goes without saying, however, that the multiverse picture raises substantial concerns both at ontological level and at the level of the semantics of set-theoretical discourse. In his *Multiversism and Concepts of Set: How Much Relativism is Acceptable?*, **Neil Barton** focuses on the semantic problems. While agreeing that the multiverse picture as advanced by Hamkins (2012) seems to square quite nicely with mathematical practice in set theory, Barton finds it problematic at a more philosophical level. He suggests that this kind of multiversism can either be interpreted as providing an ontological view, or as delivering an algebraic framework for set-theoretical practice. But analysis of both interpretations seems to leave us with a dilemma. On the one hand, the ontological picture seems to fall prey of a questionable form of relativism. Under this interpretation, each set-theoretic construction is pursued through first-order descriptions which are relative to a particular set concept defined on the background of some collection of universes, i.e., different "clouds" of universes satisfying different sentences. As a consequence, what sets exist is relative to a particular set-theoretic background, and the multiverse appears as indeterminate until a particular universe is arbitrarily chosen as a starting point. This leads, according to Barton, to an unacceptable form of relativism about reference to sets, which also has severe consequences for the indeterminacy of metalogical notions such as proof and well-formed formula. On the other hand, Hamkins' view can be interpreted algebraically, as a way of telling what is possible on any structure that satisfies the *ZFC* axioms. This seems to elude the abovementioned problems with reference, but has the unwelcome drawback of leaving Benacerrafian concerns on the nature of mathematical objects and our access to them wholly untouched, and then a significant part of our mathematical practice unexplained.

One motivation underlying the various notions of multiversism featured in Barton's discussion and Friedman and Ternullo's analysis of the Hyperuniverse is to think of them as a means of reacting, in the long run, to the threat to uniqueness of truth in set theory that resulted from Cohen's forcing technique for independence results advanced in 1963. As **Giorgio Venturi** shows in his *Forcing, Multiverse and Realism*, a thorough analysis of the notion of forcing may be required to understand several philosophical aspects of contemporary conceptions of set theory. Indeed, while forcing can be made coherent with mathematical practice by saying that what we do when we extend V is just to extend what we know about it, the idea that we can force over some countable transitive model by adding sets that still lie in V seems to suggest a robustly realist idea of a kind of set-theoretic existence which is prior to, and independent of, existence in a model. One may wonder whether this notion of existence is compatible with the notion of set as shaped by the axioms of *ZFC*, and through an historical overview Venturi answers to this question in the positive. Crucial to the understanding of forcing is a sharpening of the notion of genericity, since this is a key ingredient of forcing constructions where indeed generic extensions of a countable transitive model of *ZFC* are considered. According to Venturi, who follows Mostowski's suggestions in this, a better understanding of the notion of genericity can help disentangle several philosophical issues concerning a realist view about sets, starting from a better appreciation of the notion of arbitrary set which had a pivotal role in the development of set theory. A proper study of genericity can profit from an analysis of how alternative possible bifurcation of the set-theoretic universe led to various conceptions of the multiverse. Venturi reviews some of them, ordering them according to their different attitudes towards forms of set-theoretic realism: from platonism (in Hamkins' views, cf. Hamkins 2012), through conceptualism (in Arrigoni-Friedman's Hyperuniverse, cf. Arrigoni and Friedman 2013), to semantic realism (exemplified by Woodin's conception, cf. Woodin 2001), and finally to second-order pluralism (a view attributed to Väänänen 2014). While noticing that all these views appeal to genericity without pausing on a deeper analysis of it, Venturi eventually suggests a way of exploring the notion by taking into account sets which are generic not only with respect to some particular model, but also to some multiverse structure.

Altogether, the essays in Part II explore a variety of issues stemming from realist conceptions of such a crucial domain of mathematical objects as the one set theory attempts to characterize. Despite their diverse approaches and their specific themes, all these essays display a common underlying thread: while we seem to possess an intuitive or naïve idea of what a set is, we may end up with most variegated developments of that intuitive conception, especially on the basis of the formal theories which are employed to proceed towards more rigorous treatments. This is not something affecting only the notion of set. Attempts at a proper understanding of a variety of informal notions relevant to the philosophy of mathematics are likely to lead to rather different outcomes after formalization. As we will see, the focus on this very relationship between informal notions and formal means for rigor is shared also, and even more explicitly, by the essays in Part III.

The Logic Behind Mathematics: Proof, Truth, and Formal Analysis

Although the papers in Part III of this volume cover an apparently wide variety of topics, they are all mutually and substantially connected by the effort to clarify both the philosophical consequences and significance of formal theories and formal results, and, above all, the interaction between formalization and some crucial informal notions around which much of the research in the philosophy of mathematics and logic revolves, such as the relation between proof and truth, the relations of dependence of truths on one another, the pre-theoretical intuition of the natural number structure and how it relates to our grasping of the intended model of PA, and to the formal, metatheoretical property of categoricity of PA^2. Standing at the two opposite sides of the philosophical spectrum, formal and informal notions can positively interact or conflate dramatically, but setting sharp boundaries between the two may prove to be particularly difficult. In this respect, a rather classic area of philosophical investigation concerns to what extent our pre-theoretical understanding and inquiry of informal notions philosophically inform the formal theories and our understanding and consideration thereof, or conversely to what extent our formal theories clarify or reflect our informal or even pre-theoretical intuitions. As we will see, there is a variety of ways in which truth and formal proof-theoretical settings are at the core of the essays in this last part of the volume.

Truth and Formal Theories

Throughout the papers in this Part, the connection between truth and proof is provided by formal analysis, and still the way in which these papers explore the relations between them is rather articulate. In the first two essays, the interaction between truth and proof is directly explored via strictly formal means. The effect of this approach is that the formal analysis of the truth of the Gödel sentence $\mathcal{G}$, and of the equi-interpretability of some formal theories of truth and certain set-theoretic principles helps in precisifying and rigorizing some underlying informal notions and intuitions, such as, respectively, the meaning of $\mathcal{G}$, the intertwinement of the notions of set and truth.

In the literature, there are two main schools of interpretation of the undecidable Gödel sentence $\mathcal{G}$. According to the metatheoretical meaning (see e.g., Nagel and Newman 1958), the Gödel sentence stands for a self-referential proposition claiming unprovability of itself. In order to establish the truth of $\mathcal{G}$, one has to go look into the intended model $\mathcal{N}$, and the truth of each individual instance of $\mathcal{G}$ follows immediately. On the other hand, according to a second interpretation that attributes $\mathcal{G}$ a plainly arithmetical meaning, it is the truth of the instances of $\mathcal{G}$ that secures the truth of the Gödel sentence itself. **Mario Piazza** and **Gabriele Pulcini**'s *What's so Special About the Gödel Sentence $\mathcal{G}$?* aims at providing a further argument against the metatheoretical view and in favor of the arithmetical one. Their argument is based on Dummett (1963), and proceeds by claiming that the best way to make sense of Dummett's position is to consider that, when we prove the truth of the instances of $\mathcal{G}$, on the basis of which the truth of $\mathcal{G}$ itself is established, we have to consider generic instances of $\mathcal{G}$: namely, instances of the form $\neg Prf(n, \mathcal{G})$, where n is a generic natural number—in a slightly different parlance, one might say that n is an arbitrary natural number. This kind of proof, namely the proof of the truth of $\mathcal{G}$ carried out from the proof of the truth of its generic instances, which is also envisioned in Wright (1995), is referred to as prototypical, as in Herbrand (1931). This shift in perspective allows the authors to claim that the controversy over the epistemological priority between $\mathcal{G}$ and its numerical instances, which is the core of the metatheoretical view, endures only because the problem is ultimately ill-posed. In particular, there is no way to formally prove or disprove $\mathcal{G}$ on the basis of mathematical induction, because this would require such a proof or disproof to be carried out in PA, which cannot be the case on pain of inconsistency. Consequently, Piazza and Pulcini argue, the resort to prototype reasoning is actually unavoidable in order to achieve the truth of the Gödelian sentence. The relation of epistemic priority between $\mathcal{G}$ and its numerical instances can then be revised to the effect that, by prototypical proof, there is no need to appeal to the truth of $\mathcal{G}$ in order to recognize the truth of its individual instances.

Piazza and Pulcini argue for a sharp distinction between a substantially model-theoretical view and a proof-theoretic view in the epistemic considerations concerning the relation between the truth of the Gödel sentence and the truth of its numerical instances in PA. In contrast to this, **Carlo Nicolai**'s *More on the Systems of Truth and Predicative Comprehension*, utilizes a somewhat opposite approach based on the consideration of the equi-interpretability of typed truth theory and set-theoretical predicative comprehension in order to establish a general logico-mathematical result about the interrelation of these latter formal settings. The formal results concerning the interaction between typed truth theory and predicative set existence axioms have had a rather ample echo in several fields of research, from the foundations of mathematics, to the solutions of the semantic paradoxes, and also to a possible reduction of the ontological commitment to sets to a 'lighter' ideological commitment to notions such as truth. Nevertheless, the results available in the literature are somehow limited to those theories that take PA as the base theory. Nicolai's paper investigates a generalization of these results by treating truth and set-theoretical predicative comprehension as operations on arbitrary base theories satisfying some minimal requirements, namely being recursively enumerable. Nicolai defines three main operations: $\mathsf{T}[\cdot]$ results in a Tarskian truth theory; $\mathsf{Tp}[\cdot]$ in a typed theory of truth simulating positive inductive definitions; $\mathsf{PC}[\cdot]$ adds predicative comprehension to the base theory. In order to study $\mathsf{PC}[\cdot]$ and relate it in full generality to the truth theories also studied, a variant of it, $\mathsf{PCS}[\cdot]$, has to be taken into account: it applies to theories axiomatized by schemata in which schematic variables are replaced by second-order variables. Modulo mutual interpretability, the extension of arbitrary recursively enumerable base theories via these three operations yields equivalent results. By this general result, Nicolai shows, among other things, how set existence principles and principles governing primitive predicates for truth or satisfaction are deeply intertwined and that a general criterion of theory choice should consider them as interdependent.

Informal Notions and Formal Analysis

The last three essays of this Part and of this volume offer yet further perspectives on the general theme of the relation between truth and proof. Here, the interaction between the latter is exploited in the effort to clarify the relation between some of our informal intuitions and notions and the way in which these are systematized in formal systems. On the one hand, we find an assessment of different logical systems trying to regiment the natural thought that truths (mathematical as well as non-mathematical) stand to each other in certain relations of dependance, and that some truths are such in virtue of other more basic ones, which ground them. On the other, we see the effort of clarifying the relation between some genuinely mathematical informal intuitions, such as the intuition of the natural number structure, and model-theoretical and metatheoretical notions, such as the intended model of PA and the categoricity of PA^2. This kind of investigations may also be able to throw light on the philosophical upshot of the interaction between a somehow more informal notion of truth (as involved in the reference to the standard model of arithmetic) and the formalization of certain mathematical frameworks, thus testifying that informal intuitions may not be completely and exhaustively captured by formal settings.

In the past decade, an entirely new field of research has opened, and its implications have fast and vastly broadened. Though first analyzed systematically in the work of Bernard Bolzano, the notion of *grounding* has recently caught massive attention and worldwide interest in many different philosophical areas such as metaphysics, logic, and philosophy of mathematics. In a nutshell, the notion of grounding should capture a certain relation of priority that holds between truths or facts, and is usually signalled in the natural language by expressive devices such as 'because' or 'in virtue of,' as in 'The ball is colored because it is blue' or 'Something is the case in virtue of something else being the case.' The logical properties of this informal notion have been extensively investigated in order to formally capture both the pre-theoretical intuitions underlying it and their metaphysical consequences (cf. Correia and Schnieder 2012). On the face of this ongoing debate, grounding is either defended as both philosophically and formally substantial, or questioned as problematic (cf. Bliss and Trogdon 2014). **Francesca Poggiolesi**, in *A Critical Overview of the Most Recent Logics of Grounding*, aims at a twofold result: on the one hand, to present in a clear and faithful way two of the most recent contributions to the logic of grounding, namely Correia (2013) and Fine (2012); on the other hand, to question the formal principles describing the notion of grounding proposed by these logics. As mentioned, the notion of grounding is rather complex and has been examined from several, different perspectives, e.g., metaphysical, historical, and logical. Since much of the formal work that has been carried out in recent years is mostly interested in the logical properties of such a notion, in order to argue for her conclusion, Poggiolesi tackles grounding from a proof-theoretical point of view. According to this perspective, grounding is a proof-theoretic relation that reveals ontological hierarchies of truths. As such, though, it has to comply with many (if not all) of the properties that have been put forward for the standard calculus of natural deduction. Nevertheless, under the assumption of the proof-theoretical nature of grounding, Poggiolesi shows that this is not the case, especially with respect to negation, disjunction, and the metalogical properties of associativity and commutativity of the conjunction and the disjunction. On this basis, Poggiolesi argues that some of the formal principles that should capture the notion of grounding in Correia's and Fine's logics need to be changed and improved.

Informal notions have always played a rather substantial part in philosophical investigations: the philosophy of mathematics is no exception. In the debate on the status of formal theories of arithmetic, our informal understanding of this branch of mathematics seems to play a role as to why the intended model of the corresponding formal theories is salient with respect to nonstandard models. So, a natural question arises: what makes the intended structure of natural numbers the standard model of arithmetic? Is there any way we can explain the emergence of $\mathbb{N}$ over nonstandard interpretations? **Massimiliano Carrara**, **Enrico Martino**, and **Matteo Plebani**'s *Computability, Finiteness and the Standard Model of Arithmetic* addresses the question of how we manage to single out the natural number structure as the intended interpretation of our arithmetical language. According to Horsten's (2012) computational structuralism, the reference of our arithmetical vocabulary to $\mathbb{N}$ is determined by our knowledge of some principles of arithmetic, like those axiomatized in PA, paired with a pre-theoretical computational capability, namely a pre-theoretical ability to compute sums. Carrara, Martino, and Plebani take issue with such a view and submit an alternative answer to the question concerning the salience of the standard model of arithmetic. According to the authors, both our understanding of the axioms of PA and of how to compute sums correctly rest on something more fundamental, namely our ability to generate the relevant syntactical entities that constitute a formal theory like PA and are the basis on which the addition algorithm works. This, in turn, rests on our ability to grasp a primitive notion of finiteness. It is the intuition of this latter pre-theoretical, absolute notion of finiteness that allows the singling out of the structure of natural numbers.

While Carrara, Martino and Plebani focus on a specific proposal, i.e., Horsten's computational structuralism, concerning the salience of the intended model of PA, by rejecting it and advancing a further suggestion based on the pre-theoretical notion of absolute finiteness, **Samantha Pollock**'s *The Significance of a Categoricity Theorem for Formal Theories and Informal Beliefs* scrutinizes the role that categoricity plays in the interaction between our beliefs about informal mathematical theories (e.g., arithmetic) and the properties enjoyed by formal mathematical systems (e.g., PA^2). By offering a characterization of the requirements a theory should satisfy in order to be legitimately considered as either formal or informal, a pattern of informal notions mirrored by formal properties is recognized: for instance, we informally require that an informal mathematical theory is about a unique model or structure (e.g., the natural number structure as for arithmetic) or has an intended interpretation. These kind of informal properties are invoked in discussions on the significance of categoricity for formal mathematical theories. On this view, categoricity shows to be a two-faced property, having two kinds of philosophical significance. On the one hand, it has formal significance when it is invoked with respect to formalization: "An argument pertains to the formal significance of categoricity if it takes some informal beliefs about" an informal mathematical theory "T_I, and assesses the extent to which being categorical makes a formalization T_F of T_I adequate (i.e., faithful) with respect to those beliefs." On the other hand, it may be invoked with respect to informal mathematical theories: "An argument pertains to the informal significance of categoricity if it takes a particular formalization T_F of T_I as adequate (i.e., faithful), and assesses the extent to which its being categorical (or not) is instructive with respect to what we should informally believe about T_I." Potential consequences of this distinction arise. In particular, a potential source of circularity in Shapiro's *ante rem* mathematical structuralism (see Shapiro 1991) arises out of what appear to be arguments for both kinds of philosophical significance with respect to categoricity, thus showing that implicit claims surrounding the significance of categoricity can lead to philosophical missteps without due caution.

References

Antonelli, G. A. (2010a). Notions of invariance for abstraction principles. *Philosophia Mathematica*, *18*(3), 276–292.

Antonelli, G. A. (2010b). Numerical abstraction via the Frege quantifier. *Notre Dame Journal of Formal Logic*, *51*(2), 161–179.

Antonelli, G. A. (2013). On the general interpretation of first-order quantifiers. *Review of Symbolic Logic*, *6*(4), 637–658.

Arrigoni, T., & Friedman, S.-D. (2013). The hyperuniverse program. *Bulletin of Symbolic Logic, 19*(1), 77–96.

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy, 70*(19), 661–679.

Bliss, R., & Trogdon, K. (2014). Metaphysical grounding. *Stanford Encyclopedia of Philosophy*.

Correia, F. (2013). Logical grounds. *Review of Symbolic Logic* (1), 1–29.

Correia, F., & Schnieder, B. (2012). Grounding: An opinionated introduction. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding: Understanding the structure of reality* (pp. 1–36). Cambridge: Cambridge University Press.

Dummett, M. (1963). The philosophical significance of Gödel's theorem. *Ratio*, *5*, 140–155. Reprinted in *Truth and other enigmas*, pp. 186–201, 1978, Harvard University Press.

Fine, K. (2012). Guide to ground. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding* (pp. 37–80). Cambridge: Cambridge University Press.

Hale, B., & Wright, C. (2001). *Reason's proper study: Essays towards a neo-fregean philosophy of mathematics*. Oxford: Oxford University Press.

Hamkins, J. D. (2012). The set-theoretic multiverse. *Review of Symbolic Logic*, *5*(3), 416–449.

Herbrand, J. (1931). Sur la non-contradiction de l'arithmétique. *Journal für die reine und angewandte Mathematik*, *166*, 1–8.

Hofweber, T. (2005). Number determiners, numbers, and arithmetic. *Philosophical Review*, *114*(2), 179–225.

Horsten, L. (2012). Vom Zählen zu den Zahlen: On the relation between computation and arithmetical structuralism. *Philosophical Mathematica, 20*(3), 275–288.

Joyal, A., & Moerdijk, I. (1995). *Algebraic set theory* (Vol. 220). Cambridge: Cambridge University Press.

Lolli, G., Panza, M., & Venturi, G. (2015). *From logic to practice. Italian studies in the philosophy of mathematics*. Dordrecht: Springer.

MacLane, S. (1978). *Categories for the working mathematician* (2nd ed.). Dordrecht: Springer Verlag.

Moltmann, F. (2013). Reference to numbers in natural language. *Philosophical Studies*, *162*(3), 499–536.

Nagel, E., & Newman, J. R. (1958). *Gödel's proof*. London: Routledge.

Pantsar, M. (2014). An empirically feasible approach to the epistemology of arithmetic. *Synthese*, *191*(17), 4201–4229.

Pincock, C. (2012). *Mathematics and scientific representation*. New York: Oxford University Press.

Väänänen, J. (2014). Multiverse set theory and absolutely undecidable propositions. In Kennedy (Ed.), *Handbook of set theory* (pp. 180–208). Cambridge: Cambridge University Press.

Woodin, H. (2001). The Continuum Hypothesis. Part I. *Notices of the American Mathematical Society*, *48*(6), 567–576.

Wright, C. (1995). Intuitionists are not (Turing) machines. *Philosophia Mathematica*, *3*(1), 86–102.

In Memoriam of Aldo Antonelli

Andrew Arana and Curtis Franks

These days people are quick to take positions, “hot takes.” *ZFC* is the right foundation of mathematics, first-order logic is the true logic, the nature of mathematical knowledge is intuitionist, the axiom of projective determinism is true, nominalism is true, and so on. These are bold statements, and they draw attention. The result is the literature that moves quickly and as a result yields little change, little persuasion, little clarification, and little wisdom. It comes to seem dogmatic: one’s reputation is connected with one’s intellectual position, and thus changing your mind comes with the loss of professional status.

Aldo was not like this. Aldo’s work was about exploring new possibilities, alternatives to views that could otherwise calcify into mere dogmas. One can think of them as playing a role in an intellectual infrastructure to which spaces are regularly wrongly closed off. “that approach won’t work, for such and such a reason”: Aldo would provide technical results that opened those approaches up again. Whether you take those approaches, that’s up to you. But it’s not their impossibility that should stop you.

Mathematics has a reputation in the academy of being dry and lifeless, and among mathematical topics, logic has this reputation more than any other. Aldo didn’t see it this way. “I will take beauty over truth any day,” he would say, and while perhaps jarring at first, any seasoned mathematician will recognize the insight. Aldo, who could do just about anything, focused on logic because that found it to be incomparably beautiful. It shaped the way he taught and also the way he advised students. Discussing an amazingly talented undergraduate student with interests in logic as well as existentialism and post-Kantian themes, Aldo’s idea of

A. Arana
CRNS, IHPST (CNRS and University of Paris 1, Pantheon-Sorbonne), Paris, France
e-mail: andrew.arana@univ-paris1.fr

C. Franks
Department of Philosophy, University of Notre Dame, South Bend, USA
e-mail: cfranks@nd.edu

how to encourage her to look more into logic was to say "Just tell her that it is so beautiful, that she will love it." His sense and love of beauty fueled his distinct sense of what questions were worth asking, what questions worth pursuing. Aldo spent hours every day just working through recent papers and central texts in algebra, set theory, topology, geometry, and analysis. But he never wrote about these things. Aldo knew, not everything in modern mathematics, but close to everything that was beautiful.

We'll continue with a brief discussion of Aldo's early work. He worked on the foundations of defeasible consequence and non-monotonic logics, of Quine's New Foundations set theory, and of non-well-founded set theory. Rather than try to summarize the results, we can see this work as expressing the order of Aldo's approach to problems. One might say that there were some technical gimmicks to work out, and he had the relevant tools to discover and present these things. But this is wrong. Aldo was interested in the results that were the most beautiful; the ones that unpacked the most hidden connections, the ones that made us rethink the greatest number of our inhibiting preconceptions. Because he didn't care much about, and possibly didn't even understand, ideas about which set theories were correct, he was able to feel his way to the mathematical relationships that disclosed the greatest number of such insights.

Aldo applied this method with particular focus to logicism and in particular to the thesis that arithmetic is logic. Frege's logicist program had the following two stages: (1) to define numbers as extensions of concepts, and (2) to derive logically the theorems of arithmetic from that definition. That these two stages cannot both be carried out is more or less consensus today. Where there is consensus, though, there is need for light: for what possibilities are being closed out by the consensus? We want to talk about three projects of Aldo that opened such possibilities for logicism.

We'll begin with "Frege's New Science," written with Robert May. Aldo broke the question of whether Frege's logic could carry out metatheory into two parts: (1) can metalogical questions about Frege's logic be posed in Frege's logic itself? and (2) is metatheory necessarily model theory, in which one varies the meanings of propositions in order to prove for instance independence results? Aldo and Robert answered "no" to (2), in light of Frege's insistance that one cannot reinterpret the meanings (references) of nonlogical terms of axioms, since axioms express thoughts. They argue that Frege saw, if not particularly clearly, a way to develop metatheory in which one replaces the reinterpretability of meanings with a kind of permutability of nonlogical vocabulary on which no meanings are changed. While the question of what vocabulary is nonlogical arises, Aldo and Robert sketch an argument (at least nearly) available to the Fregean that this question too can be handled by a permutability argument. Thus this article opens space for a new approach to understanding Frege's metatheory by pursuing a technical development.

Next, let's turn to "Frege's Other Program," also written with Robert May. Here Aldo and Robert explore a different possibility for reckoning the two logicist stages of defining numbers as extensions of concepts, and of deriving logically the theorems of arithmetic from that definition. As Aldo and Robert put it, one can attempt "to show in a nonlogical theory of extensions, where numbers are concepts, not objects, that Peano Arithmetic can be derived." In doing so they identify a non-logicist but still broadly Fregean program for deriving arithmetic. This program clarifies the causes of the contradiction entailed by Basic Law V, providing a new counterexample to Hume's Principle. Here again new spaces are opened, by investigating an extensional theory of arithmetic without supposing Hume's Principle. One can then consider to what extent such a theory could be judged a vindication of logicism.

Next, we'll turn to Aldo's 2010 article "The Nature and Purpose of Numbers," notable not least because it appeared in one of American philosophy's top journals, the *Journal of Philosophy*. This article investigates a possible version of Fregean logicism, one that differs from other developments in that it does not reduce arithmetic to set theory. Aldo takes cardinal properties of the natural numbers as the starting point, and derives structural properties of the natural numbers from them, rather than the other way around as is done in typical set-theoretic reductions of arithmetic. Aldo's idea is that "in keeping with the broadest and most general construal of logicism, cardinality notions… deal directly with properties and relations of concepts—rather than matters of existence of objects such as numbers—[and thus] cardinality notions properly can be regarded as having a logical character." They are logical notions, Aldo argues, because relations of concepts are quantifiers: more precisely, they are, Aldo argues, generalized quantifiers. The ordinary existential and universal quantifiers can be thought of as relations of concepts: the existential quantifier as the collection of all nonempty subsets of the domain of quantification, and the universal quantifier as the collection of all subsets of the domain that contain the domain as their only member. But these are not the only two relations that yield quantifiers, on the theory of generalized quantifiers, and in particular one can consider the Frege quantifier, which holds between two concepts F and G when there are no more Fs than Gs. In fact, these are all first-order quantifiers, Aldo argues. He then shows how the Frege quantifier can be taken as logically basic, and shows how one can derive the basic features of the natural numbers from such a logic.

We'd like to note in particular Aldo's way of characterizing his accomplishment in this article. "Accordingly, we take the logicist claim that cardinality is a logical notion at face value, and rather than arguing for it (perhaps by providing a reduction to some other principle), we set out to explore its consequences by introducing cardinality, in the form of the Frege quantifier, as the main building block in the language of arithmetic."

This passage illustrates beautifully the idea that Aldo's approach to problems was prescind from particular theoretical stances, and to explore the consequences of these stances. With a clearer understanding of these consequences, of their fruits, one can better evaluate the costs and benefits of particular positions. The job of the logician is to explore these consequences. And Aldo was a logician.

In this 2010 article, Aldo also noted how the Frege quantifier, a first-order quantifier, can be given a generalized Henkin interpretation. In Henkin's work, one restricts the range of quantification to just a subset of the power set of the full domain. Models for second-order logic can then be specified by giving both the domain and a universe of relations over the domain. Aldo's novel insight was that such an interpretation can also be given for first-order quantifiers. Within the context of Fregean logicism, this permits one to prescind from concerns about whether second-order logic is logic (another one of those dogmatic debates).

His 2013 article in the *Review of Symbolic Logic*, "On the general interpretation of first-order quantifiers," expanded on this novel insight. 60 years or so after Henkin's groundbreaking work on generalized models, Aldo observed what no one else ever noticed, namely, that the notion of a generalized model can be formulated already for first-order languages. The irony is sharp: The "first-order case" of the Henkin construction becomes an extension (not a restriction) of the familiar second-order case, where the notion of models given by filters over the full power set construction is more intuitive. This is, in our opinion, Aldo's deepest work.

In an article entitled "Life on the Range: Quine's Thesis and Semantic Indeterminacy," published this summer, Aldo pursued the consequences of this technical development for the evaluation of Quine's dictum that to be is to be the value of a bound variable. As Aldo notes, this dictum flows from Quine's view that second-order logic is "set theory in sheep's clothing." Since second-order logic on Quine's view has ontological commitments, it is not really logic. Aldo observes that his work on generalized models puts pressure on Quine's views. Since the first-order quantifiers can be interpreted to be extensions of second-order quantifiers, the ontological commitments of second-order logic are also ontological commitments of first-order quantifiers. In the closing sentence of this article, Aldo writes that "this last realization can contribute to the establishment of second-order logic on the same safe footing as first-order logic."

In closing, it seems to us that most professional philosophers spend more time advancing their own research programs than they spend learning. This strikes us as completely unreasonable, and we are pretty sure that our attitude derives from Aldo's influence. We ask ourselves: Why would I, or anyone really, care more about what I have to say than about what some 20 or so brilliant historical figures have already said? Do I love logic, math, and philosophy, or do I love professional credits? Everyone in our world initially loved the former, and it is a disgrace, Aldo taught, to abandon this idea. And, he taught, if you persevere in your love for the most beautiful ideas in mathematics, philosophy, and logic, your own contributions will trickle in at the right time. Those ideas will not come close to being the most interesting things you have to talk about. But they will not only be true, they will be beautiful.

Chapter 1
Mathematics in Philosophy, Philosophy in Mathematics: Three Case Studies

Stewart Shapiro

Abstract The interaction between philosophy and mathematics has a long and well articulated history. The purpose of this note is to sketch three historical case studies that highlight and further illustrate some details concerning the relationship between the two: the interplay between mathematical and philosophical methods in ancient Greek thought; vagueness and the relation between mathematical logic and ordinary language; and the study of the notion of continuity.

Keywords Greek mathematics · Modern logic · Vagueness · Continuity

Almost from the beginning, philosophy has had a fascination with mathematics. In (very) rough terms, rationalism is the attempt to apply the methodology of mathematics to all discursive thinking, to science and philosophy in particular. The main thesis of empiricism, the main opposition to rationalism, is that all concepts and all factual knowledge are derived from sensory experience. Apparently, mathematics does not fit that mold; it at least appears to be an a priori enterprise. Empiricists took this seriously, and went to great lengths to accommodate mathematics.

The purpose of this note is to sketch three historical case studies that highlight and further illustrate some detail concerning the relationship between mathematics and philosophy. There is traffic in both directions.

It is not much of a distortion to claim that both philosophy and mathematics were born in Ancient Greece. Greek mathematicians were fascinated with problems like doubling the cube and squaring the circle, problems whose resolution went beyond anything practical. It was not a matter of getting close to a doubled cube or squared circle—close enough for whatever practical purposes one may encounter. That much is an easy exercise, not one that would baffle generations of brilliant mathematicians. The Greeks had their sights on mathematical precision, an exactitude that goes well beyond anything observable or even anything measurable.

S. Shapiro (✉)
Department of Philosophy, The Ohio State University, Columbus, OH, USA
e-mail: shapiro.4@osu.edu

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_1

So we begin in ancient Greece, and, in particular, with Plato, since that body of work sets the tone for two millennia of interaction between mathematics and philosophy. Plato's admiration of the accomplishments of Greek mathematicians is abundantly clear. Some recent scholars have focused attention on the influence of the development of mathematics on Plato's philosophy. We see a sharp and dramatic contrast between Plato and his teacher Socrates.

From what we can tell, mostly from Plato's writings, Socrates' main interests were in ethics and politics, not in mathematics and not in what passed for science during that period. We are familiar with stories of Socrates discussing justice and virtue with anyone who would listen and talk. At his trial, he declared (*Apology* 38):

> …I tell you that to let no day pass without discussing goodness and all the other subjects about which you hear me talking and examining both myself and others is really the best thing a man can do, and that life without this sort of examination is not worth living.

Socrates typically proceeds by eliciting the beliefs of an interlocutor and then, through careful questioning, drawing out surprising and unwanted consequences of those beliefs. The interlocutor is thus challenged to reexamine his beliefs.

Socratic method, then, is a technique for weeding out false beliefs. If it does produce truth, it is only by a fumbling process of elimination or trial and error. After a series of revisions, one may end up with a set of beliefs that pass the challenge of rigorous examination. Clearly, Socratic method does not result in certainty. Upon examination, we learn that some of our beliefs are false or confused, but the method does not inevitably point to which of the beliefs are false or confused. The method is thus fallible and hypothetical. Socrates took his greatest wisdom to be his belief that he does not know.

The methodology of the mature Plato does not resemble that of Socrates, in any of these ways. Gregory Vlastos (1991, 107) argues that the differences are due, in large part, to the influence, and accomplishments of ancient mathematics.

Plato notes in passing that mathematics is "universally useful in all crafts and in every form of knowledge and intellectual operation—the first thing everyone has to learn" (*Republic* 523). Of course, this is true today. Think of how many disciplines have extensive mathematics prerequisites. Moreover, by Plato's day, and, indeed, in our day, one needs intense and prolonged study to master mathematics—even if we restrict ourselves to the mathematics that finds application in the sciences. Thus, Plato held that one needs intense and prolonged study for any "form of knowledge and intellectual operation". For Plato, that includes philosophy.

Unlike Socrates, then, Plato held that philosophy is not for everyone. In the city-state envisioned in the *Republic*, only a few carefully selected leaders, the Guardians, are to engage in philosophical reflection. The vast majority of the inhabitants of the envisioned state are told to get their direction from these leaders and to mind their own business; that is, to mind only their own business: farmers stick to farming; cooks stick to cooking; physicians stick to healing. Like farming, cooking, and healing, philosophy is left to the experts—those experts being the chosen Guardians. So, against his celebrated teacher, Plato insisted that for the vast majority of people, the unexamined life is well worth living, or at least for these masses, the unexamined life must be led; there is no alternative. If Plato had his way, the examined life would be forbidden to almost everyone, to all but the most worthy.

For Plato, a full decade of the Guardians' training is devoted to mathematics. They do little else from age 20 to age 30. Plato's reason for this is that, in order to rule well, the Guardians need to turn their focus from the world of Becoming—the physical world that our bodies inhabit—to the world of Being, where Truth lies. According to Plato, mathematics "draws the soul from the world of change to reality". It "naturally awakens the power of thought…to draw us towards reality" (*Republic* 521). Of course, this is only for the few souls capable of such ascent, and this ascent is essential to being a decent Guardian.

In sum, then, for Plato, mathematics is an essential preliminary to philosophical study and it demands a long period of intense study. No wonder that most of us have to live our lives in ignorance of True Reality, and must rely on Guardians for direction as to how to live well. Plato's break with the expansive views of his teacher is thus understandable, if not admirable.

Plato's fascination with mathematics may also be responsible for his distaste with the hypothetical and fallible Socratic method. Mathematics proceeds (or ought to proceed) via proof, not a mere random trial and error that may or not end up with true knowledge. Mathematics at least appears to produce necessary truths, and it has an exquisite certainty in its results. As Plato matures, Socratic method is gradually supplanted. In the *Meno*, Plato uses geometric knowledge, and geometric demonstration, as the paradigm for all knowledge, including moral knowledge and metaphysics. In that dialogue, Plato wants to make a point about ethics, and our knowledge of ethics, and he explicitly draws an analogy with geometrical knowledge. It is not much a distortion to see Plato as a forerunner to rationalism. (See Chap. 3 of Shapiro (2000), and the references cited there.)

Today, of course, there are those that accept at least some of Plato's main theses, and there are those who oppose them. We have our platonists and our anti-platonists. There are also elitists who see philosophy as a discipline only for the few who are property trained, and there are those who see philosophy as for everyone, keeping the Socratic theme of the unexamined life alive (although there does not seem to be much correlation between the platonists and the elitists). This is not the place to settle these disputes. We turn to our second case study.

Logic is another subject that began in ancient Greece. The first logician, of course, was Aristotle, Plato's main student and main opponent. Aristotle, too, admired mathematics, and accommodated the accomplishments of mathematics into his philosophy. However, he did not follow his teacher and make mathematics the model for all advances in knowledge. But that is a topic for another day, to be pursued by someone more versed in the history.

In the *Prior Analytics*, one can discern at least a precursor to a formal language, along with a sort of model-theoretic semantics and, notably, a deductive system (see Corcoran 1972, 1974). It is limited to a rather simple conglomeration of syllogisms; all propositions are in the form All A is B, No A is B, Some A is B, or some A is not B. A contemporary logician might dismiss this Aristotelian system as little more than an example, perhaps one fit for teaching, but not holding much in the way of theoretical interest. It is too simple. The ancient world did see other logical frameworks, whose languages had different, sometimes greater, expressive resources. There was notable logical activity throughout the medieval period, but most of that focused on relatively minor extensions of the Aristotelian categorical system. In one of the greatest (and most embarrassing) ironies, Immanuel Kant declared that "since the time of Aristotle [logic] has not had to go a single step backwards… What is further remarkable about logic is that until now it has not been able to take a single step forward, and therefore seems to all appearances to be finished and complete" (*Critique of pure reason*, B viii).

Of course, Kant could not have be more wrong about this. Logic exploded, to breathtaking heights, once mathematicians got involved and, indeed, took over. It began in the final decades of the nineteenth century, with algebraists like Boole and Schröder, Peirce, and DeMorgan, and continued through the likes of Frege and Hilbert, and then Tarski, Church, Kleene, Rosser, Henkin, Gödel et al. The list includes some of the finest mathematical minds of the past 200 years or so.

Today, logic is taught in both mathematics and philosophy (and computer science) departments, and there is considerable overlap in the courses, at least up to a point. There is proof theory, model theory, set theory, recursive function theory, and a host of other sub-disciplines.

When I was younger, I was fond of proclaiming that mathematics provides a good case study for many philosophical issues. Mathematical languages, for example, are ripe for examining theories of reference, since messy matters of causality, observation, and the like do not get in the way. Mathematical knowledge provides a good paradigm for semantics and epistemology, or at least some semantics and epistemology. Figure out how mathematics works, and we have a start on understanding knowledge generally, or at least one kind of (presumably a priori) knowledge.

I still think this, but the brash claims must be tempered. The focus on mathematics brings distorting influences as well, even for logic. And this takes us back to philosophy. In rough terms, logic is science of correct reasoning, or at least correct deductive reasoning. The founders of contemporary mathematical logic—say Frege, Hilbert, Gödel, and Church—were focused exclusively on reasoning in mathematics itself, and on forms of inference in mathematical languages. Like Plato, they were simply uninterested in vagueness, ambiguity, context-dependence and a host of other messy features of natural languages and, thus, with features of ordinary correct (even deductive) reasoning about the messy world we inhabit. A staple of model-theoretic semantics is that in an interpretation of the (formal) language, each predicate gets a sharp extension, a set of objects to which the predicate applies. There is no room in the system for borderline cases, nor for predicates whose extensions grow or shrink as we learn more about the world and encounter unexpected cases. There is no room for what Waismann (1945) calls the "open texture" of language and thought. Predicates in formal logics have fixed, sharp, and complete extensions—independent of context. One cannot just take it for granted that ordinary language, and ordinary deductive reasoning about ordinary things, will fit smoothly into the mold of model theory.

A large part of contemporary philosophy, and contemporary semantics in linguistics, consists of attempts to extend the success of mathematical logic to ordinary language. We see Montague semantics, possible-world semantics for modality and propositional attitudes, including knowledge and belief, many-valued and supervaluationist accounts of vagueness, fuzzy logic, temporal logic,… It is an exciting enterprise.

I am reminded of the child who is given a hammer for her birthday. As she goes around the house, she discovers, to her delight (and her parents consternation), that everything she encounters is in need of hammering. We philosophers were bequeathed a powerful batch of tools: model theory, proof theory, set theory, recursion theory, you name it. It is natural for us to hold that everything we encounter will yield to those tools, and be enlightened thereby.

So perhaps my youthful enthusiasm was a bit too optimistic. Perhaps we should keep in mind that there may be important mismatches between mathematical languages, especially how they are regimented in logic, and the rough and ready world of ordinary thought or even scientific thought. It may be that things are not as determinate and clear as the mathematics for which our cherished tools were developed.

I suggest that epistemicism concerning vagueness is an rather extreme example of the phenomenon I have in mind. A small, but dedicated and articulate group of philosophers maintain that all legitimate predicates of natural language, such as “bald”, “tall”, “about noon”, and “heap” have, in fact, sharp extensions (e.g., Sorensen 2001; Williamson 1994). It is just that we do not and cannot know where these boundaries lie. There is a single hair-molecule that makes the difference between bald and hirsute, a single height that makes the difference between tall and not-tall. Indeed, the same even goes for predicates that seem designed to be vague: a single nanosecond separates the time for which it is “about noon” from the times when it is not. I suspect that the main attraction of such views is that, according to them, ordinary, model-theoretic semantics applies to natural language. To put it the other way, epistemicism is all but inevitable if one insists that model-theoretic semantics gives the correct account of how natural language functions.

In sum, we have sharp tools and we try to use them to theorize about a loose and messy subject. My own, broadly contextualist account of vagueness (Shapiro 2006) is an attempt at a sort of compromise. The main philosophical thesis is that when it comes to borderline cases of vague predicates, speakers can to either way, without compromising their linguistic competence and without sinning against any non-linguistic facts. I call this “open-texture”.

The study contains a model theory designed to capture the idea. It is a variant on the Kripke semantics for intuitionistic logic (and the modal logic S4). The main difference is that at each node—at each world—each predicate has an extension and a disjoint anti-extension. Also, the compositional “semantics” goes via a notion called “forcing”, not the usual, local notion of “truth at a node/world” (and so we do not end up with intuitionistic logic). The idea is that each node represents one way that vague terms can be deployed in a conversation, consistent with the meaning of the words and the non-linguistic facts.

As some critics have pointed out, the framework contains elements that do not exactly match the way that vague terms are used in ordinary language. At least in the main version, there is a sharp cutoff between cases of which a given predicate holds of determinately and those where open-texture kicks in.

The underlying idea (developed in Chap. 2 of Shapiro 2006, see also Cook 2002) is that the framework provides a mathematical model of the way that vague terms are deployed. It represents some features of the use while idealizing on and ignoring others. There is an extensive literature on mathematical models, and it would be too much of a distraction to go further here.

Now we come to our third and final case study. I have lately gotten interested in the notion of continuity, which constitutes another example of the interaction between mathematics and philosophy. Here, too, we begin in ancient Greece. Zeno presented a number of paradoxes, aimed at showing that motion is impossible. The most fascinating and compelling of those involve the relationship between a continuous entity, such as a line segment or a period of time, and individual points (of zero size), or moments (of zero duration), thereof. Consider, for example, the paradox of Achilles and the tortoise. Suppose that Achilles and a tortoise are about to have a race. Since Achilles can run, say, a hundred times faster than the tortoise, he gives the tortoise a head start. They begin at time t0. Before Achilles can catch the tortoise, he must run to the place from which the tortoise began. Suppose he reaches that point at time t1. After that, Achilles must get to the place where the tortoise is at t1. Suppose he reaches that place at time t2. But then he must get to where the tortoise is at t2, getting there at time t3. And so on. So, it seems, Achilles cannot catch the tortoise, since before he can do so, he must complete an infinite number of runs, one to t1, a second to t2, etc.

Some ancient thinkers, the atomists, held that a seemingly continuous entity, such as a line, is composed of minimal segments, each of finite length. There simply are no points, at least physically, and perhaps even conceptually. The same goes for time: each temporal interval is composed of minimal segments, each of finite duration.

Aristotle did not follow the atomists, presumably because he took the advances in geometry seriously. His answer to at least these Zeno-paradoxes is that a line does not consist of a set of points, and a temporal interval does not consist of a set of moments, or “nows”, each of zero length. Points are only the extremities of line segments; moments are just the extremities of temporal events. Contra the atomists, any given line segment, no matter how small, can be bisected, thus producing an even smaller segment. A bisection produces a point in the middle of the segment. The same goes for periods of time. But, according to Aristotle, we cannot say that the point, or the moment, existed before the bisection. The infinity of points on a line segment, or in an interval of time, is only potential, not actual. There are infinitely many places and times where Achilles could have stopped, and if does stop at one of them, he’ll produce a segment of time and a segment of space and, with those, a point and a “now”. But if he does not stop, then there are no such points and moments to worry about. Achilles does not have to pass through points and moments when chasing the tortoise. Similarly, a given line segment can be bisected as many times as one wishes. There is no upper limit to how many points we can produce. But there is never an actual infinity of points on the line segment.

For Aristotle, a continuous entity, such as a line, is viscous, in the sense that it is held together as a kind of unity. One consequence of this, it is thought, is that it is impossible to cleanly divide a continuous entity into disjoint pieces. When a line segment is split, for example, two new endpoints are created, one for each of the segments. The contemporary idea of open and closed regions is foreign to this approach. Does it make sense wonder whether, say, a block of wood is open or closed on its boundary?

Aristotle defines a substance to be continuous, as opposed to merely contiguous, if when two items of that substance are placed together, the extremities between them are absorbed into each other. So water, wet mud, and wet dough are all continuous, blocks of wood are not.

The matter of continuity occupied medieval philosophers, and gets mixed, in interesting ways, with (western) theology. Some of the medieval, Christian logicians, such as Burley and William of Ockham, worried about Aristotle's notion of potential infinity. For us finite humans, it is not possible to bisect a line segment to infinity, and so for us, the points are only potential. For God, however, the points are actual. He can "see"—at once—all of the places where a line could be cut, i.e., he can "see" all of the points. Thus, what is potential for us humans is somehow actual to God (see Normore 1982).

The contemporary, now orthodox classical conception of the continuum does not follow Aristotle (and the human side of the medievals). It is hard for the contemporary mathematician and logician to make sense of the notion of a merely potential infinity. For Georg Cantor, every potential infinity must have an actual infinity underlying it:

> I cannot ascribe any being to the indefinite, the variable, the improper infinite in whatever form they appear, because they are nothing but either relational concepts or merely subjective representations or intuitions (imaginationes), but never adequate ideas (Cantor 1883, 205, note 3).
>
> ... every potential infinite, if it is to be applicable in a rigorous mathematical way, presupposes an actual infinite (Cantor 1887, pp. 410–411).

This is the now-orthodox perspective. A line, for example, just is a set of points. Zeno's paradoxes are addressed in other ways, by paying attention to countable and uncountable infinities, and formulating more clearly how and when infinitely many finite quantities can be summed.

The contemporary continuum is not viscous: a line segment can be divided cleanly into two, three, or as many pieces as one wants—even uncountably-many point-sized pieces—with nothing left over or created along the way.

I think it generally agreed that one key impetus for this development was the scientific need for discontinuous functions, to model discontinuous motion and change, in a continuous background. The relationship between the mathematics and science here is a long and complex story, one that has yet to be written.

To connect our second and third case studies, some twentieth and twenty-first century mathematicians have shown how to develop a broadly Aristotelian account of the continuum—including the rejection of actual infinity, in a rigorous fashion. The most natural ways to do this, however, require a break with (so-called) classical logic. In particular, the law of excluded middle is given up. The logic is intuitionistic. To put it in more metaphysical mode, one must give up the thought that for any given object x, say a point, and every given property P, either P determinately holds of x or P determinately fails to hold of x. Indeed, once this logical-cum-metaphysical matter is absorbed, the contemporary Aristotelian can have his cake and eat it, too: a line does consist of points, and yet it is viscous (although Aristotle himself seemed to be committed to excluded middle, at least in this context). Another item of note is that, in the intuitionistic theories, a continuous entity, such as a line segment, cannot be broken cleanly.

But, unlike the Aristotelian conception, any attempt to break a line segment, for example, leaves something out. A typical metaphor is that when a line is cut, something sticks to the knife. For Aristotle, in contrast, something new is created, namely the new endpoints. See Bell (1998), and, for example, Troelstra and van Dalen (1988).

It should be noted that the intuitionistic theories do not have an intermediate value theorem. Although this theorem is quite intuitive, it does not comport well with the theme that all functions are smooth, or at least continuous—a theme that is tied to the viscosity theme broached above.

Geoffrey Hellman and I have developed (and are developing) another point-free account of the continuous. Like Aristotle, our theories do not recognize the existence of points, at least not as parts of regions. And we use so-called classical logic. Our first account (Hellman and Shapiro 2012, 2013) is limited to one dimension, and makes essential use of actual infinity (and so the account is only "semi-Aristotelian"). It can be shown to be equivalent to the now orthodox Dedekind-Cantor account, by defining points as certain limits. There is also a more Aristotelian theory (Hellman and Shapiro forthcoming), one that eschews the use of actual infinity, and there is work-in-progress to extend the treatment to two and higher-dimensions, and to various non-Euclidean theories.

The best conclusion, I submit, is that there simply is no monolithic notion of the continuous, or at least none that is consistent. There are a number of intuitive principles that underlie the intuitive notion, but it seems that these principles are inconsistent with each other. The notion is better understood as one that is up for sharpening, in various incompatible ways, rather than a straightforward analysis.

This concludes our case studies. The treatments were overly brief and, for that reason, a bit simplified. But I trust that they amply highlight the rich interplay between mathematics and philosophy.

References

Aristotle. (1941). In McKeon, R. (Ed.), *The basic works of Aristotle*. New York: Random House.

Bell, J. (1998). *A primer of infinitesimal analysis*. Cambridge: Cambridge University Press.

Cantor, G. (1883). *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*. *Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen*. Leipzig, Teubner.

Cantor, G. (1887). Mitteilugen zur Lehre vom Transfiniten 1, 2. *Zeitschrift für Philosophie und philosophische Kritik*, *91*(1987), 81–125 and 252–270, *92*(1988), 240–265 (reprinted in Cantor (1932), 378–439).

Cantor, G. (1932). In Zermelo, E. (Ed.), *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Berlin: Springer.

Cook, R. T. (2002). Vagueness and mathematical precision. *Mind*, *111*, 225–247.

Corcoran, J. (1972). Completeness of an ancient logic. *Journal of Symbolic Logic*, *37*, 696–702.

Corcoran, J. (1974). Aristotle's natural deduction system. In J. Corcoran (Ed.), *Ancient logic and its modern interpretations* (pp. 85–131). Dordrecht, Holland: D. Reidel.

Hellman, G., & Shapiro, S. (2012). Towards a point-free account of the continuous. *Iyyun*, *61*, 263–287.

Hellman, G., & Shapiro, S. (2013). The classical continuum without points. *Review of Symbolic Logic*, *6*, 488–512.

Hellman, G., & Shapiro, S. (forthcoming). An Aristotelian continuum. *Philosophia Mathematics*, *3*.

Kant, I. (1998). *Critique of pure reason*, P. Guyer & A. W. Wood (Eds. & Trans.). Cambridge: Cambridge University Press.

Normore, C. (1982). Walter Burley on continuity. In N. Kretzmann (Ed.), *Infinity and continuity in ancient and medieval thought* (pp. 258–269). Ithaca: Cornell University Press.

Plato, (1945). *The Republic of Plato*, (F. Cornford, Trans.). Oxford: Oxford University Press.

Plato, (1961). In Hamilton, E. & Cairns, H. (Eds.), *Collected Dialogues* (Vol. 71). Bollingen Series Princeton: Princeton University Press.

Shapiro, S. (2000). *Thinking about mathematics: the philosophy of mathematics*. Oxford: Oxford University Press.

Shapiro, S. (2006). *Vagueness in context*. Oxford: Oxford University Press.

Sorensen, R. (2001). *Vagueness and contradiction*. Oxford: Oxford University Press.

Troelstra, A. S., & van Dalen, D. (1988). *Constructivism in mathematics* (Vols. 1 & 2). Amsterdam: North Holland Press.

Vlastos, G. (1991). *Socrates: Ironist and moral philosopher*. Ithaca, New York: Cornell University Press.

Waismann, F. (1945). Verifiability. In *Proceedings of the Aristotelian society*, Supplementary Volume, 19, pp. 119–150 (reprinted in Flew, A. (Ed.) (1968). *Logic and language* (pp. 117–144). Oxford: Basil Blackwell.

Williamson, T. (1994). *Vagueness*. New York, London: Routledge.

Part I
The Ways of Mathematical Objectivity: Semantics and Knowledge

Chapter 2
Semantic Nominalism: How I Learned to Stop Worrying and Love Universals

G. Aldo Antonelli

Abstract Aldo Antonelli offers a novel view on abstraction principles in order to solve a traditional tension between different requirements: that the claims of science be taken at face value, even when involving putative reference to mathematical entities; and that referents of mathematical terms are identified and their possible relations to other objects specified. In his view, abstraction principles provide representatives for equivalence classes of second-order entities that are available provided the first- and second-order domains are in the equilibrium dictated by the abstraction principles, and whose choice is otherwise unconstrained. Abstract entities are the referents of abstraction terms: such referents are to an extent indeterminate, but we can still quantify over them, predicate identity or non-identity, etc. Our knowledge of them is limited, but still substantial: we know whatever has to be true no matter how the representatives are chosen, i.e., what is true in all models of the corresponding abstraction principles. This view is backed up by an "austere" conception of universals, according to which these are first-order objects, i.e., ways of collecting first-order objects. Antonelli thus claims that second-order logic does not import any novel ontological commitment beyond ontology of first-order, naturalistically acceptable, objects. Moreover, in the case of arithmetic, even if Hume's Principle is construed as described above, Frege's Theorem goes through unaffected, since there is nothing in its proof that depends on an account of the "true nature" of numbers. A viable construal of logicism is then given by the combination of semantic nominalism and a naturalistic conception of abstraction. [*Editors note*]

Keywords Abstraction principles and Frege's Theorem · Invariance and logicality · Arithmetic with Frege quantifiers · Semantic nominalism

---

I am grateful to Tyrus Fisher and Elaine Landry for valuable comments and criticism.

G.A. Antonelli (✉)
University of California, Davis, CA, USA

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_2

2.1 Introduction

When it comes to the question of the existence and proper status of *abstract entities* (such as numbers, propositions, functions, and even sets) philosophers have found that turning to science can often lead to conflicting conclusions. On the one hand, scientifically inspired empiricism poses the urgent question of our epistemic access to such abstract entities, given their non-spatio-temporally located and causally inert nature. The question was famously posed by Benacerraf (1973) and again by Field (1989): if all knowledge is at least partially—but ultimately—based on experience, then the specific nature of these abstract entities places them beyond the reach of human cognition. This is true at least on a causal theory of knowledge, as Hale points out in Hale (2001, p. 169), but even if somehow the causal theory is rejected it remains true that abstract entities, *as such*, and as ordinarily conceived, remain outside the purview of *physical* science.

On the other hand, scientific naturalism insists that the claims of science be taken at face value, even when they involve putative reference to the abstract entities of mathematics: scientific practice cannot be challenged on purely philosophical grounds. As Lewis (1991, p. 59) famously put it:

> I am moved to laughter at the thought of how *presumptuous* it would be to reject mathematics for philosophical reasons. How would *you* like the job of telling the mathematicians that they must change their ways, and abjure countless errors, now that *philosophy* has discovered that there are no classes?

Taking the statements of mathematical or natural science "at face value" requires, among other things, that we ascribe to them the logical form that they appear to have, that we construe the singular terms thus identified as referential, and supply the resulting sentences with Tarskian truth conditions. In particular, when scientific statements contain singular terms purportedly referring to abstract entities—numbers, sets, functions, etc.—the requirement enjoins us to identify appropriate referents for such terms and specify the relations in which such objects can stand to other objects, abstract or ordinary as they might be.

Lewis is right in pointing out the philosophers' *hubris* in dictating the mathematicians' ontology. But even if we adopt the naturalistic viewpoint, we still have to provide philosophically compelling answers to the questions pressed on us by Burgess and Rosen (2005):

(a) *The ontological question*: What are abstract entities?
(b) *The epistemological question*: How can we have knowledge of them?

Answers to the two questions are often inter-related in interesting ways, and sometimes, as noticed, even in tension with one another. Our aim in what follows is to resolve the tension by articulating an account that is naturalistic in taking the claims of mathematical science at face value, while at the same time addressing empiricist reservations about the special status of abstract entities.

2.2 Fregean Platonism

Platonism about abstract entities is, roughly speaking, the view that such entities are endowed with independent existence (independent, that is, of the cognitive processes of the human mind). While this form of realism is attractive for its ability to account for the role that abstract entities play in our scientific practice, a tension arises when trying to articulate a concurrent epistemological view. But Wright (1983) offered supporters of the realist approach a new and promising avenue for the introduction of abstract entities: one could view such entities as delivered by *abstraction principles*, with the consequence that one needed only to be conversant with an already familiar range of concepts in order competently to employ singular terms referring to such entities and to recognize truths about them.

Second-order abstraction principles take the form:

$$\alpha(F) = \alpha(G) \text{ if and only if } \mathscr{R}(F, G),$$

asserting that the α-abstract of F is the same as the α-abstract of G if and only if the two concepts F and G are related by the equivalence relation $\mathscr{R}$. The abstraction is second-order because it involves the assignment of abstracts to *concepts*. On the other hand, first-order abstraction principles assign abstracts to first-level objects (whatever those may be); such principles can be represented schematically as

$$\delta(x) = \delta(y) \text{ if and only if } R(x, y).$$

First-order abstraction principles provide an important contrast class, since they exhibit properties that are significantly different from the second-order case (more about which later). The main example of higher-order abstraction that will concern us here is the same one that was put forward by Wright as paradigmatic: Hume's Principle, introducing *numbers* as second-order abstracts of the equivalence relation $F \approx G$, which obtains when just as many objects fall under F as they fall under G. If Hume's Principle is indeed to allow a *non-problematic* way of introducing numbers, "just as many" is, crucially, to be spelled out without, in turn, appealing to the notion of number (using the second-order statement to the effect that there is a bijection between the F's and the G's).

Rosen is quite adamant about the advantages of introducing abstract entities, and numbers in particular, by means of abstraction principles. Unreflective forms of Platonism suffer from the "vaguely feudal" picture of two distinct realms of objects, the one inhabited by ordinary objects of everyday life, and one that is home to a "vast catalogue of abstract objects": on this basis, "[P]latonism can only sound insane" (Rosen 1993, p. 152). On the other hand, Fregean Platonism "promises to undermine the hopeless two-world-picture by identifying facts about abstract objects with facts about ordinary concrete things". This identification is in line with another feature of ordinary thought, the so-called "supervenience of the abstract", the idea that facts about abstract objects are fixed by the facts about concrete objects. To use Rosen's example: although *King Lear*, the play, is an abstract object distinct from each and all its concrete manifestations (books, performances, etc.), the facts about the play are determined by the array of concrete objects in the world. When applied, *mutatis mutandis*, to the case of mathematics, this insight gives a form of Platonism "without tears", which is

> the only serious attempt […] to establish [P]latonism while restoring the intelligibility of the supervenience of the abstract and thereby diffusing the pernicious effects of the [two-world] picture (Rosen 1993, p. 154).

Consider now the first-order abstraction principle for directions, introduced by Frege in §65 of *Grundlagen* (Frege 1980, p. 76), which is also extensively discussed by Rosen:

$$\mathsf{dir}(a) = \mathsf{dir}(b) \text{ if and only if } a \parallel b.$$

Fregean Platonists argue that the abstract entities thus introduced, *directions*, are completely unproblematic, in that anyone who already grasps the notion of parallelism of two lines will also, *eo ipso*, grasp the notion of direction. This claim is made precise by showing that the adjunction of the principle to a theory of parallelism is *conservative* over the base theory of parallelism, in that it does not allow the derivation of new facts expressible in the base theory. There is a sense in which adjoining the abstraction principle for directions to the base theory does not really tell us anything about line parallelism that we did not know before.

The abstraction principle for directions shares, in fact, such unproblematic nature with all other first-order abstraction principles. All such principles are conservative over the corresponding base theory. But to take first-order abstraction principles as representative of the larger class of second-order principles is both misleading and instructive. It is misleading because not all abstraction principles are as innocent as first-order ones (as Frege himself came to realize), but interesting since, if we accept the theoretical surplus that can accompany higher-order abstraction principles, then there is a lot to be learned from the comparison.

2.3 Semantic Nominalism

At the other end of the spectrum from unreflective Platonism, we find *eliminative nominalism*. Although nominalism dates back to medieval scholastics, its modern form originates with Goodman and Quine (1947), and it has among its contemporary proponents Field (1980) and Hellman (1989). Eliminative nominalism is engaged in a *reductive* program, advocating the elimination of abstract entities from scientific discourse, very much in the same way in which Weierstraß's (1854) arithmetization of analysis eliminated infinitesimals from the calculus. Such an elimination, when achieved, would then blunt standard *indispensability arguments* for abstract entities (first explicitly proposed by Putnam 1971). Such arguments are based on the notion that acceptance of a scientific theory carries along ontological commitment to all (and only?) those entities that are indispensable to it, but nominalization—if successful—would show that abstract entities are not indispensable.

It is fair to say that this form of reductive eliminativism fails, for various reasons, to achieve its goal in full generality, a case that has been made by Colyvan (2001) among others (but see also Panza and Sereni 2013). Colyvan in particular argues that it is not sufficient that abstract entities be purged from a scientific theory in order for us to drop our commitment to such entities; in addition one also needs to show that the resulting theory is *better* than the one it replaces along the usual metrics employed by scientists in the assessment of theories (simplicity, elegance, explanatory power, etc.). And this is a tall order that no form of eliminative nominalism has been able to fulfill.

Fregean Platonism, in particular, presents a particular challenge for the nominalist project, in that abstract terms introduced via abstraction principles are not eliminable. This is in fact the force of the so-called *Caesar's problem*: abstraction principles allow us to settle the truth value of identities between abstract terms by reducing them to expressions no longer containing reference to *abstracta* ("the number of F = the number of G" can be replaced by the second-order statement saying that there is a bijection of the F's onto the G's); but the principle is silent when it comes to identities between an abstract term and a singular term such as "Julius Caesar". In this case, Fregean Platonism can offer no recipe for the elimination of abstract terms. (Thus, the Caesar problem, far from being a shortcoming of the Platonistic strategy, offers a vigorous defence against attempts at nominalization).

If nominalism is to stand a chance *vis à vis* the naturalistic force of indispensability arguments, an important distinction is needed. Until now, we have grouped together the nominalistic rejection of abstract entities of all sorts: numbers, propositions, classes, etc. But in fact these entities are quite different from one another, and versions of nominalism differ according to which class of entities is targeted. Let us introduce a distinction between two different kinds of nominalism:

- *A-Nominalism* addresses the opposition between abstract and concrete objects.
- *U-Nominalism* addresses the opposition between universals and particulars.

A-Nominalism is only concerned with what kinds of *objects* there are, i.e., whether among the objects populating our ontology there are any that are *abstract* in the sense that they are non-spatio-temporally located and causally inert. One needs to be careful here not to collapse the abstract/concrete distinction onto the theoretical/observable distinction. Obviously objects that are spatio-temporally located need not be directly observable, as long as they are a kind of objects posited by natural science. But perhaps even more importantly, in line with Russell's mature logical atomism (Russell 1918–1919), there is no absolute notion as to what counts as an *object*. By "object" we understand any kind of entity that is *saturated* (in Frege's sense), an entity, that is, of the kind suitable for being the referents of a singular term. Pragmatic choices in language design and progress in conceptual analysis might, over time, change the range of entities that we regard as "saturated" in this sense.

A-Nominalists as characterized here deny that there are abstract objects; since it follows that all objects are spatio-temporally located, all objects are therefore firmly within the purview of natural science. A priori methods apply, of course, but any a priori methods as to the existence and properties of objects have to survive the special kind of push-back that only experience can provide. This stance is fully in line with a naturalistic conception in that it puts science first when it comes to questions regarding the existence and properties of objects.

U-Nominalism is concerned with the distinctions between particulars and universals, between what is traditionally characterized as a *this* and what is *one over many* instead. In Fregean terms, this is the distinction between saturated and unsaturated entities. For our purposes we will adopt a rather austere view of universals: we identify universals with entities in the type hierarchy over some domain D of objects; in particular collections of such entities or maps between such entities. Universals so conceived correspond to unsaturated entities in Fregean terms: predicates, or (more generally) functions. The view is austere because it attends to just one of the various functions of universals, foregoing for instance any consideration of their intensional nature, but it is sufficient for our purpose of developing a naturalistic account of abstract entities.

If we had to choose a label for the kind of account to be proposed, the choice would fall on "Semantic Nominalism", as the doctrine that advocates A-Nominalism but embraces universals under the austere conception. The term is introduced by Hale and Wright (2001, p. 352) for a view that they ultimately (but, we argue, prematurely) reject. Quine (1951) proposes a distinction between *ontology* and what Quine refers to, for lack of a better term, as *ideology*. Ontology is a doctrine as to what kind of *objects* are to be countenanced, whereas ideology deals with the same question within the realm of universals. We want to propose an account that is ontologically nominalistic but ideologically liberal, albeit in an "austere" way.

This is not the place to give a defense of A-Nominalism, as the empiricist motivations for such a position are well known, and a rejection of A-Nominalism in favor of unreflective Platonism would be, as Rosen pointed out, "insane". We will instead motivate a rejection of U-Nominalism by drawing on *semantic* and *logical* considerations. Lest the realists get too excited, though, we anticipate that the argument in the next section will only support a "thin" notion of universals, one which falls way short of any attempt at reification.

2.4 The Semantic Case for Universals

Humans are, to some extent and to varying degrees, conversant with the kind of austere universals identified above. We recognize them, apprehend them, reason about them, and use them to express facts about first-order objects; also, although not nearly as often as philosophers like to think, we occasionally explicitly assert their existence (a distinction between *expressing* and asserting the existence of a given higher-order entity can be found in Antonelli and May 2012). Much of the evidence for this claim is linguistic. Prior (1971) was perhaps the first to point out that ordinary language is replete with instances of quantification over austere universals of higher type; a similar case is also found in Higginbotham (1998, p. 3), who considers the sentence:

He's everything we wanted him to be,

which contains a quantifier over austere universals, viz., second-order properties. In a similar vein, Prior (1971, p. 48) proposes the following examples of *non-nominal quantification* (i.e., quantification over expressions of a syntactic category other than names):

I hurt him somehow;
He's something I am not—kind.

These sentences are obtained by existential generalization over adverbial and noun phrases, respectively (so that "somehow" represents a *third-order* quantifier):

I hurt him by treading on his toes;
He is kind, but I am not.

The claim that humans are conversant with austere universals and other higher-order entities needs to be qualified in at least two ways. The first, obviously, is that our facility with universals need not extend all the way up the type hierarchy: there is no sense in which humans have native proficiency with logics of order ω, and in fact natively such proficiency extends at most to the first few levels of the hierarchy. But even restricting ourselves to the first few levels—and this is the second qualification—we need to recognize that any "access" to such higher-order entities may (and probably must) be limited to some collection that falls short of the full domain of the appropriate type. For instance, when contemplating properties of objects, we implicitly restrict our attention to some salient collection of them, much in the same way in which when using ordinary first-order quantifiers we restrict their range to some implicitly given salient collection of objects.

We claim that such access is real access to austere universals nonetheless, witness the fact that it is not restricted to generic semantic proficiency with the expressions in question, but is in fact specific and detailed enough to provide validation of some characteristic inference patterns. Consider first the familiar case of ordinary inferences involving quantifiers other than the ones from the first-order predicate calculus:

Most of the students went to the lecture;
Most of the students passed the course;
Therefore, some student who went to the lecture passed the course.

A moment's reflection reveals, even to the untrained mind, that the inference is valid. However, a rigorous proof of this fact, while not difficult, is not completely trivial, as it requires some facility with the combinatorics of finite maps (in particular, the proof exceeds the inferential power of first-order logic, as the quantifier "most" is not first-order definable, see Peters and Westerståhl 2006). It is precisely this kind of proficiency with unsaturated entities (injective maps between finite sets) that bears witness to our apprehension of austere universals. Notice that no claim is made here as to the precise nature of this proficiency: it might well be that recognition of the validity of the argument is based on some geometric intuition, or it might be based on some rudimental formal system. Be that as it may, the upshot is that elementary facts about higher-order entities, even of some complexity, underpin our inferential ability.

Similar inferences are also available further up the type hierarchy; the following is adapted from Higginbotham (1998, p. 3):

All we expected him to be was honest, polite, and scholarly;
He is mostly what we expected him to be;
Therefore, he is either honest or polite.

Here again we see some elementary combinatorics involving higher-order entities, and many more examples could be found.

We should mention one more aspect of the dependence of our semantic proficiency upon access to higher-order entities. As argued in Antonelli (2013, 2015), the semantics of first-order quantifiers, including, but not limited to, the ordinary quantifiers "All" and "Some", has a higher-order dimension. When considered from the point of view of the theory of generalized quantifiers (and as adumbrated in Frege's *Grundgesetze*), a first-order quantifier is a third-order predicate, i.e., a predicate of predicates. According to this view, then, the existential quantifier $\exists$ is a predicate applying to all and only the non-empty subsets of the first-order domain D. But the quantifiers need not be interpreted as predicates over the full power set of the first-order domain, and we can allow first-order models with a non-standard second-order domain. If this is the case, then the semantics of first-order quantifiers is not really determinate until and unless a (possibly non-standard) second-order domain is specified alongside the first-order one. The technical details are given in Antonelli (2013), and some of the philosophical repercussions are explored in Antonelli (2015). But for now let us observe that while this is yet more evidence for the role played by higher-order entities, one particular consequence of the framework is that it allows for a clear analysis of issues of ontological commitment. In a recent article Florio and Linnebo (forthcoming) argue that it is only within the context of non-standard, i.e., Henkin models for higher-order logic that the question of ontological commitment can sensibly be asked: the same holds true of non-standard *first-order* models, of course, a fact that is not in the foreground until such models are given due consideration.

The remarks in this section thus have a bearing on the issue of the proper demarcation of logic and mathematics. The idea that at least parts of mathematical knowledge can be traced back to logic is not new, and in fact it figures prominently in several versions of logicism and neo-logicism. But rather than focusing on the logicist tradition originating with Frege, it is important to notice that human proficiency with universals as we have characterized them is front and center already in Dedekind's version of logicism:

> If we scrutinize closely what is done in counting an aggregate or number of things, we are led to consider the ability of the mind to relate things to things, to let a thing correspond to a thing or to represent a thing by a thing, an ability without which no thinking is possible. (Dedekind 1988, p. 14)

Thus the ability to "let a thing correspond to a thing," is singled out by Dedekind as a fundamental function of the human mind, and this is an ability that is grounded in a facility with austere universals. Almost ninety years on, a similar point is made by Feferman:

> …when explaining the general notion of *structure* …we implicitly *presume as understood* the ideas of *operation* and *collection* …at each step we must make use of the unstructured notions of operation and collection to explain the structured notions to be studied. The *logical* and *psychological priority* if not primacy of the notions of operation and collection is thus evident. (Feferman 1977, p. 150)

Both Dedekind and Feferman emphasize the role that our ability to engage entities corresponding to the unsaturated parts of language plays in the development of mathematics, and on the basis of the semantic evidence provided, this seems right. But this does not settle the issue of which side of the boundary between logic and mathematics such ability is to be located: where does logic end and mathematics start? Those who question the very idea of such a boundary, or would like to move the boundary much further out from where it is usually found (as it is for many versions of logicism), are happy to say that at least a large part of mathematics just *is* (or is reducible to) logic.

The conception of austere universals, on the other hand, supports the claims that the ability to "let a thing correspond to a thing" and to collect such things in various ways, is indeed to be properly located within logic. On this view, second- and higher-order logic is, indeed, *logic*. But it is important to recognize at the same time that this conception of higher-order entities is ontologically *thin*: engagement with the universals does not add anything—any *thing*—to our ontology. This is a point forcefully made by the supporters of Plural Logic, a position that Florio and Linnebo (forthcoming) refer to as "Plural Innocence". But on our view there is nothing special discriminating plural quantifiers from *bona fide* higher-order quantifiers: if the former are ontologically innocent, so are the latter. *Pure* second-order logic is indeed an ontologically weak theory: all the second-order validities are true in the standard second-order model whose first-order domain comprises one and only one object (just like in the first-order case). It is this conception of universals as ethereal insubstantial entities (the "Will-o'-the-wisp" theory of universals, as it were) that underpins our logical abilities.

However, the option of *reification* of the universals is always open, and, in many ways, tempting. Universals are reified when they are identified, or made to correspond to, objects in the domain. For instance, this is the framework of modern set theory, in which first-order objects are related by $\in$, so as to mimic the way in which first- or higher-order entities "fall under" universals. There is of course no underestimating how immensely successful this enterprise has turned out to be. But it should also be clear that this kind of reification leads out of the domain of pure logic, and into the mathematical domain.

Not all attempts at reification are successful, however. Frege's Basic Law V is exactly one example: we cannot, on pain of inconsistency, assign objects to what Frege called first-level concepts (i.e., second-order universals) in such a way that concepts under which the same objects fall are assigned the same object. In order to escape the inconsistency one needs either to restrict the range of concepts to which objects are assigned, or to coarsen the equivalence between concepts, as is done, for instance, in Hume's Principle. This second strategy figures prominently in the next section.

2.5 The Abstraction Mystique

Fregean Platonism, which according to Rosen holds the best hope for a realistic account of abstract entities, including those of mathematics, rests on two components: the mathematical core of the view, supplemented by an extra-mathematical thesis concerning the special status of entities introduced by abstraction (in the case of arithmetic, by Hume's Principle).

The mathematical core of the view concerns the basic fact that positing an abstraction principle of the form:

$$\alpha(F) = \alpha(G) \text{ if and only if } \mathscr{R}(F, G),$$

simply asserts the existence of a mapping α from the second-order domain of Fregean concepts back into the first-order domain, in such a way that concepts related by $\mathscr{R}$ are mapped onto the same object. The crucial issue is the *coarseness* of $\mathscr{R}$, i.e., the size of the collection of the equivalence classes of concepts induced by $\mathscr{R}$. The coarser the relation $\mathscr{R}$, the easier it is to satisfy the principle. At one extreme we find the coarsest, i.e., the universal relation on the second-order domain: the induced abstraction operator maps all concepts onto one and the same object, and the corresponding abstraction principle is always satisfiable as long as the first-order domain is non-empty. At the other end of the spectrum is the finest relation, i.e., the identity, whose induced operator assigns to each concept a distinct object from the first-order domain. This is the relation employed in Frege's Basic Law V, and it is unsatisfiable over standard domains (i.e., second-order domains comprising the full power set of the corresponding first-order domain). Somewhere in between we find Hume's Principle, where the relation $\mathscr{R}$ holds between concepts F and G precisely when just as many objects fall under F as they fall under G. This is where interesting mathematics can take place: Hume's Principle allows the recovery of second-order arithmetic, provided that second-order logic is assumed in the background, a result that is justly celebrated as "Frege's Theorem". Moreover, as we assign objects to concepts as their *numbers*, i.e., as representatives of their respective equivalence classes, we obtain as a consequence that numbers themselves can be *counted*. This mathematical fact lies at the heart of the abstractionist approach to arithmetic: since there are $n + 1$ numbers less than or equal to n, Hume's Principle cannot be satisfied over finite domains.

Superimposed to the mathematical core is the *abstraction mystique*, i.e., the idea that abstraction principles are the preferred vehicle for the delivery of a special kind of objects—abstract entities, and in the case of Hume's Principle, *numbers*—which would be otherwise unattainable given their causal inefficacy and non-spatio-temporal location. This is in fact the point raised by Rosen (1993), when he characterizes Fregean Platonism as the best hope for mathematical realists, precisely because abstraction principles appear elegantly to dispose of the tension between the ontological problem and the epistemological problem raised by abstract entities, giving us a picture of such entities that is not "insane". But it is important to see that the mathematical core and the accompanying mystique are in fact independent, and that one can put the former to good use without embracing the latter.

Let us recall once again that mathematically, a second-order abstraction principle is just a functional mapping of second-order entities into the first-order domain respecting a given equivalence relation. In order to obtain the most general characterization of such mappings, non-standard second-order domains have to be allowed, i.e., domains that fall short of the full power-set of the first-order domain. Accordingly, by a *model* we understand a pair (D_1, D_2), where D_1 is any non-empty set and D_2 a collection of subsets of D_1 (the model also provides interpretations of the appropriate type for the extra-logical constants of the language). In practice, several closure conditions of the second-order domain are relevant, for instance:

- $D_2 = \mathscr{P}(D_1)$ (i.e., D_2 is standard);
- D_2 contains all definable subsets of D_1 (in some given background language $\mathscr{L}$).
- D_2 is closed under some class Π of permutations of D_1.

The last closure condition comes down to the following: given a permutation π of D_1 and a subset $X \in D_2$, the point-wise image of X under π, denoted by $\pi[X]$, is also in D_2. This is important, since closure under permutations is one of the features characterizing logical notions. Tarski (1986) indeed proposes this kind of invariance as a necessary and sufficient condition for logicality (a claim that has been variously disputed, see, e.g., Bonnay 2008); but all we need is the much weaker claim that invariance is a necessary condition for some higher-order notion to claim a logical character. Antonelli (2010b) introduces several different notions of invariance that might be taken to be germane for abstraction principles:

- Invariance of the equivalence *relation* $\mathscr{R}$;
- Invariance of the *operator* α;
- Invariance of the abstraction *principle*.

We say that the equivalence relation $\mathscr{R}$ is invariant if and only if for all $X \in D_2$, we have $\mathscr{R}(X, \pi[X])$ (assuming D_2 is closed under permutations). Similarly we say that the abstraction operator α is invariant if and only it commutes with π, i.e., $\pi(\alpha(X)) = \alpha(\pi[X])$. The invariance of $\mathscr{R}$ is indeed a desirable property, but it does not speak to the logical status of α. For abstraction to be correctly characterized as a logical operation, α would have to be invariant. However, the invariance of $\mathscr{R}$ is mostly incompatible with the invariance of α, and to make matters worse, no operator α satisfying Hume's principle is invariant (see Antonelli 2010b, propp. 6 and 7).

But abstraction is a form of reification, so it should not come as a surprise, given our discussion in the previous section, that it turns out to be mathematical rather than logical. However, the challenge remains for empirically-minded philosophers to make sense of abstraction in terms that do not postulate a separate realm of entities (regardless of whether they are logical or mathematical in character). The way out of the difficulty lies in the realization that reification does not by itself commit one to the acceptance of a separate realm. Reification as embodied by abstraction principles consists only in the positing of a correspondence of the appropriate kind between second-order entities and first-order objects. There is no concomitant assumption about the ultimate nature of those objects, no accompanying mystique. The objects targeted by the reification are drawn from the same first-order domain D constituted by the objects of natural science. When those objects are delivered as referents of abstract terms, all we need to know about them is that they function as representatives of equivalence classes of second-order concepts.

The characteristic mathematical role played by abstraction principles resides in laying down constraints on the cardinality of the second-order domain D_2 (which is allowed to be non-standard) *vis-à-vis* the cardinality of the first-order domain. The finer the equivalence relation $\mathscr{R}$ the more upward pressure is exerted on the cardinality of D_1, to the point that where $\mathscr{R}$ is maximally fine (i.e., when $\mathscr{R}$ is identity of extensionally given concepts), the corresponding abstraction principle is not satisfiable save on highly non-standard second-order domains, as Cantor's theorem tells us. But doesn't naturalism enjoin us to take the claims of (mathematical or natural) science at face value? And those claims very clearly regard abstract entities as first-order objects, as when we say that the number of the planets is eight, or that the speed of light equals $2.99792458 * 10^8$ meters per second. On this view, the issue of the special nature of *abstracta* simply does not arise—abstracta are simply ordinary objects recruited for a specific mathematical purpose.

One question that arises on the naturalistic view of abstraction concerns the advantages, if any, of postulating Hume's Principle as opposed to one of the many available versions of the Axiom of Infinity. For if the proper function of Hume's Principle is not to deliver a special class of entities, but only to guarantee that there are enough of them for the proof of Frege's Theorem to go through, why shouldn't we just postulate an appropriate Axiom of Infinity (to the effect that, say, a given binary relation R is serial, irreflexive, and transitive)? But the Axiom of Infinity does not provide for a way to identify, among the countably many objects, those that function as *numbers*, and it does not supply the means to *refer* to them, *quantify* over them, and *predicate* various properties of them (including identity and difference). And it is not even clear how to perform all of these linguistics tasks without accessing resources (set-theoretic or otherwise) that are not germane to arithmetic proper. In contrast, Hume's Principle delivers all of this in one fell swoop: not only does it force the domain to be infinite, by identifying representatives of equivalence classes of concepts, but it also allows us to refer to such representatives, *numbers*, while remaining within a properly arithmetical context.

When it come to matters of ontology, naturalism, properly understood, is the view that ontological commitment should extend to all entities required for the truth of scientific claims—*but no further*. It is this "no further" clause that naturalists tend to forget. In the case where scientific claims make reference to abstract entities, i.e., entities such as numbers that are the referents of abstract terms, all that naturalism requires is that we have a way of appropriately selecting first-order representatives for equivalence classes of higher-order entities. This is possible in the case of claims of (pure or applied) arithmetic as long as we have enough first-order entities to make such an assignment possible. Thus in such cases naturalism can be reconciled with the empiricist worldview. True, the naturalist is still committed by Hume's Principle to the existence of countably many first-order objects; but this is a commitment shared with the Fregean Platonist, who is thus no better off.

Burgess and Rosen (2005) go to great length to lay out the naturalistic claims faced by the empirically or nominalistically inclined philosopher. These claims can be summarized as follows:

(a) Standard mathematics is rich in theorems asserting the existence of mathematical objects.
(b) Existence theorems are accepted not just by *mathematical* standards, but by broader *scientific* standards (no empirical refutation).
(c) Philosophy cannot override mathematical or scientific standards of acceptability (this is the essence of naturalism).
(d) Hence, we are justified in believing, along with mathematicians and scientists, in the existence of the various mathematical objects.

Supporters of Semantic Nominalism can gladly embrace all of this claims, at least in the case of arithmetic, but also in all cases where the relevant mathematical objects can be introduced by means of abstraction principles.

2.6 Arithmetic with the Frege Quantifier

This section provides an outline of a formal theory of arithmetic, AFQ, which employs a cardinality quantifier F (the Frege quantifier) and a term-forming abstraction operator, Num, taking formulas as arguments (both devices are understood to bind a variable). AFQ provides an example of *non-reductive logicism*, i.e., a theory in which the standard of logicism is borne by the cardinality quantifier (expressing a logical notion, based on the arguments of Sect. 2.4). The theory still employs an abstraction principle (a form of Hume's Principle), but such a principle is characterized as extra-logical and mathematical. The details can be found in Antonelli (2010c), and more of the philosophical background in Antonelli (2010a).

Given some standard supply of extra-logical constants, the language of our theory comprises the two variable-binding devices already mentioned: the binary Frege quantifier, allowing for formulas of the form $\mathsf{F}x(\varphi(x), \psi(x))$, and the abstraction operator allowing for terms of the form $\mathsf{Num}\, x\, \varphi(x)$. The intended interpretation is that $\mathsf{F}x(\varphi(x), \psi(x))$ expresses (but does not assert) the fact that there are no more φ's than ψ's, and $\mathsf{Num}\, x\, \varphi(x)$ represents the number of the φ's.

A standard model $\mathfrak{M}$ for the language provides, just like for ordinary first-order logic, a non-empty domain D_1 of objects and interpretations for the extra-logical constants (predicate and function symbols) as well a function η from the power-set of D_1 into D_1 (η interprets the extra-logical abstraction operator). Satisfaction of a formula φ in $\mathfrak{M}$ by an assignment s to the variables is recursively defined as usual, with the two clauses:

- $\mathfrak{M} \models \mathsf{F}x(\varphi(x), \psi(x))[s]$ if and only the cardinality of $[\![\varphi]\!]_s^x$ is less than or equal to that of $[\![\psi]\!]_s^x$;
- $[\![\mathsf{Num}\, x\, \varphi(x)]\!]_s$ is defined to be $\eta([\![\varphi]\!]_s^x)$,

where $[\![\varphi]\!]_s^x$ is the extension of $\varphi(x)$ relative to s. Notice that several notions are expressible in the language, beginning with the ordinary quantifiers $\exists$ and $\forall$: $\exists x\varphi(x)$ can be represented by denying that the cardinality of (the extension of) φ is less than or equal to that of the empty set; and dually $\forall x\varphi(x)$ can be represented by the assertion that the cardinality of $\neg\varphi$ is less than or equal to that of the empty set. The expressive power of this language exceeds that of ordinary first-order logic, since we can say, for instance, that the extension of φ is Dedekind-finite by denying that there is a y such that the cardinality of φ is less than or equal to the cardinality of $\varphi(x) \wedge x \neq y$. It is convenient to introduce the abbreviations $\mathsf{Fin}\, x\, \varphi(x)$ for the latter statement, and $\mathsf{I}\, x\, (\varphi(x), \psi(x))$ for:

$$\mathsf{F}x\, (\varphi(x), \psi(x)) \wedge \mathsf{F}x\, (\psi(x), \varphi(x)),$$

which (by the Schröder-Bernstein Theorem) expresses that φ and ψ are equinumerous, i.e., they have the same cardinality. The symbol I is ordinarily used to represent Härtig's quantifier.

We now come to the extra-logical axioms of AFQ. The extra-logical constants comprise a 2-place relation $<$ and a 1-place predicate $\mathbb{N}$. A first group of axioms are definitional, they give us uniqueness conditions for various notions. Unsurprisingly, the first one is a version of Hume's Principle:

$$\mathsf{Num}\, x\, \varphi(x) = \mathsf{Num}\, x\, \psi(x) \leftrightarrow \mathsf{I}x\, (\varphi(x), \psi(x)). \qquad \textbf{Ax1}$$

The second axiom gives us a definition of $\leq$ ($x \leq y$ abbreviates $x < y \vee x = y$):

$$\mathsf{Num}\, x\, \varphi(x) \leq \mathsf{Num}\, x\, \psi(x) \leftrightarrow \mathsf{F}x\, (\varphi(x), \psi(x)). \qquad \textbf{Ax2}$$

The next axiom gives the definition of successor, modeled on Frege's original definition. The number of the ψ's is the successor of the number of the φ's if and only if there is a ψ such that there are just as many φ's as there are ψ's other than it:

$$\mathsf{Succ}\, x\, (\varphi(x), \psi(x)) \leftrightarrow \exists x(\psi(x) \wedge \mathsf{I}y(\varphi(y), \psi(y) \wedge y \neq x)). \qquad \textbf{Ax3}$$

The next two axioms are substantial, as opposed to purely definitional. The first one gives an implicit definition of the collection of natural numbers:

$$\forall y(\mathbb{N}(y) \leftrightarrow \mathsf{Fin}\, x(\mathbb{N}(x) \wedge x < y) \wedge y = \mathsf{Num}\, x\, (\mathbb{N}(x) \wedge x < y)). \qquad \textbf{Ax4}$$

The last axiom is a form of comprehension: for formulas $\theta(x, y)$, $\varphi(x)$ and $\psi(y)$, let $\theta : \varphi \xrightarrow{1-1} \psi$ say that θ defines an injection from the φ's into the ψ's:

$$\begin{aligned}\forall x\big[\varphi(x) \to \exists y\big(\psi(y) \wedge \theta(x, y) \wedge \\ &\forall z(\psi(z) \wedge \theta(x, z) \to z = y) \wedge \\ &\forall z(\varphi(z) \wedge \theta(z, y) \to z = x)\big)\big].\end{aligned}$$

Our last axiom then says that if there is a definable injection of the φ's into the ψ's then the cardinality of the former is less than or equal to that of the latter:

$$[\theta : \varphi \xrightarrow{1-1} \psi] \to \mathsf{F}\, x\, (\varphi(x), \psi(x)). \qquad \textbf{Ax5}$$

Ax5 is, of course, valid over standard models, so it is not needed as long we restrict our attention to such models. But Antonelli (2010c) also introduces a notion of *general model* for the language with the Frege quantifier, and then the role of this axiom in the proof of the Principle of Induction becomes apparent (see Antonelli and May 2012 for details).

The five axiom schemas of AFQ are enough to interpret first-order Peano Arithmetic (and perhaps more). In particular, one can prove the analogues of the Peano-Dedekind axioms:

(i) $\mathbb{N}(0)$: where 0 abbreviates $\mathsf{Num}\, x\, (x \neq x)$; i.e., 0 is a number.
(ii) For numbers p and q let $\mathsf{Succ}(p, q)$ abbreviate $\mathsf{Succ}(\mathbb{N}(x) \wedge x < p, \mathbb{N}(x) \wedge x < q)$; then every number has a unique successor.
(iii) Every number other than 0 is a successor (importantly this is provable without induction).
(iv) Succ is an injective function.
(v) $\big[\varphi(0) \wedge \forall n(\varphi(n) \to \varphi(n+1))\big] \to \forall n \varphi(n)$ (requires **Ax5** over general models).

In particular, since Succ is an injection of $\mathbb{N}$ into itself, the theory proves that $\mathbb{N}$ is not Dedekind-finite.

2.7 A Naturalistic View of Abstraction

According to *Naturalism*, as we understand it here, the purview of natural science encompasses all there is—the furniture of the universe from the micro level to cosmological scale. Accordingly, there is nothing, no *thing*, that falls outside its purview. As a corollary, the philosopher has to accept the existence of any entities that are required for the proper functioning of science; this is the thrust of indispensability arguments. But an equally important, if somewhat less emphasized, corollary is that our ontological commitments should *not* extend beyond what is so required.

Our austere universals do not contravene to this second corollary, for they are not—*ex hypothesi*—first-order objects on a par with all the other first-order objects comprising the domain of natural science. Rather, they are ways to collect such objects according to this or that principle, or (in Dedekind's words) to "represent a thing by a thing". Various forms of reification of our universals do contravene the second corollary, for instance when issues are raised about their location or causal efficacy. But at least in some cases, i.e., those in which reification takes place through the positing of abstraction principles (arithmetic being the most prominent example), the resulting framework is compatible with the naturalistic outlook. Abstraction does take us outside of the domain of pure logic and deep into mathematical territory, but it does not do so by positing entities that fall outside the purview of science. This is not to imply, of course, that (say) numbers *as such* are spatio-temporally located or causally efficacious. On the naturalistic view of abstraction, numbers are whatever entities are selected as representatives of equivalence classes of concepts, and as we know the selection can be carried out in infinitely many ways, as long as we assume the existence of an infinite number of objects (there can be no arithmetic, failing this assumption). *Abstracta* are delivered as referents of abstraction terms, but no further claims are made as to their special nature. In this sense, if the eliminative nominalism of Field or Hellman is analogous with Weierstraß' elimination of infinitesimals, naturalistic abstraction is more in line with Hilbert's introduction of *ideal elements*.

The view combining naturalistic abstraction with semantic nominalism represents a natural equilibrium point between the competing demands of realism and anti-realism: naturalistic abstraction shares the advantages of Fregean Platonism, but without the accompanying mystique. In particular it meets the following requirements, which are commonly presented as the features that make a realistic outlook on abstract entities desirable:

(a) The left-hand sides of abstraction principles is taken at face value; there is no need to re-interpret identities between abstract terms as other than statements of objectual identity.
(b) The denotations of terms so introduced are available for quantification, identification (and differentiation), predication, and reference. In particular, numbers themselves can be counted, which is at the heart of the abstractionist account of arithmetic.
(c) No *additional* commitment to abstract entities is necessitated by the positing of abstraction principles *as such*—ontology is not to be settled through semantics, and mathematics is not logic—as long as the concomitant cardinality requirements are met.

This view, or something very similar, can be found at various junctions in the literature, only to be dismissed (with great effort) by both nominalists and realists. For instance, a similar position is taken up for consideration by Rosen, who points out that the proposed interpretation works for first-order abstraction:

> Whenever the Fregean introduces new singular terms by abstraction on an equivalence relation among ordinary particulars, it will be possible to interpret the new terms as referring to these ordinary things. (Rosen 1993, p. 174)

But the strategy fails at the higher-order, essentially because of the non-conservative nature of second-order abstraction, the primary example being Frege's doomed Basic Law V. We know of course that the failure of Basic Law V does not extend to Hume's Principle (provided we are ready to admit the existence of infinitely many objects), and even Basic Law V is satisfiable over non-standard domains.

For another example of how the strategy naturally arises, only to be dismissed, here is Dummett's assessment of the case of first-order abstraction:

> The stipulation that the direction of line *a* is to be the same as that of a line *b* just in case *a* is parallel to *b does not determine whether the direction of a line is itself a line or something quite different*: this contextual definition indeed has a solution, but it is far from unique. Even if the requirement were to be made that every direction should itself be a line, the stipulation would in no way determine which line any given direction was to be; it could, in fact, be any line whatever [...] The contextual definition might well be defended on the ground that we do not need to know anything about directions save what it tells us: as long as we know that the direction of *a* is the same as that of *b* just in case *a* is parallel to *b*, *we are quite indifferent* to what, specifically, the direction of *a* may be, or any other facts about it. Frege makes it plain in §66 [of *Grundlagen*] that this defence would not satisfy him at all ...(Dummett 1991, p. 126, emphases added)

The view entertained by Dummett in the above passage is very similar to the naturalistic conception of abstract entities. Once we have the notion of parallelism between two lines, we can pick any representatives of the equivalence classes—or indeed any distinct objects whatsoever—to play the role of directions. This is completely unproblematic in the case of first-order abstraction, for first-order abstraction principles are always satisfiable just in the way indicated by Dummett, and in fact they are conservative over the underlying theory (since any model of the underlying theory can be expanded to a model of the principles). The reason Frege would not embrace the definition, Dummett continues, is that:

> It is an inexcusable defect in a proposed definition of the direction-operator that it fails to tell us what, specifically, the direction of a given line is to be; and hence it must be replaced by an explicit definition that does tell us that. (Dummett 1991, *ibid.*)

Needless to say, the mathematics is unaffected by this "inexcusable defect," and neither are we. Things are somewhat different in the case of second-order abstraction (which is not conservative over the background theory), but the intuition has to remain the same. Abstraction principles provide representatives for equivalence classes of second-order entities; such a choice of representatives is possible when, and only when, there are no more equivalence classes in the second-order domain (which might be non-standard) than there are first-order objects. Such representatives will be available provided the first- and second-order domains are in the equilibrium dictated by the abstraction principles. But otherwise the choice of representatives is unconstrained. The proof of Frege's Theorem goes through unaffected on this construal of Hume's Principle, since *formally* there is nothing in the proof that depends on an account of the "true nature" of numbers.

We are now well positioned to solve the ontological problem, by answering the question, *What are abstract entities?*: They are the referents of abstraction terms. Such a referent is to an extent indeterminate, but we can still work with such terms, quantify over their referents, predicate identity or non-identity, etc. What about the epistemological question, *How do we know about abstract entities?* Our knowledge of them is limited, but still substantial. In particular, we know whatever has to be true no matter how the representatives are chosen, i.e., what is true in all models of the corresponding abstraction principles. We won't know anything about the special nature of the representatives—their spatio-temporal location, for instance. But we will know whatever follows from the positing of such representatives.

Before we proceed to the conclusion, we consider a particular objection to semantic nominalism developed by Hale and Wright, which deals with issue of *trans-world abstraction* (see Hale and Wright 2001, pp. 352ff). The argument then runs as follows. On the assumption that particular lines are chosen as directions:

(a) The direction of a line exists in all worlds where the line exists (abstract terms are referential).
(b) Assuming a domain of contingents, there is a world w where line a exists orientationally unchanged at the same time as all other lines (parallel to a in the real world) change their direction.
(c) Therefore at w, and hence at all worlds, $\mathsf{dir}(a) = a$ (a is the only possible representative of its equivalence class).
(d) Thus in the real world, given that abstract terms are rigid, $\mathsf{dir}(a) = a \neq b = \mathsf{dir}(b)$ even when $a \parallel b$.

Observe that the argument requires that the "representatives" be selected from their respective equivalence classes, while this is not in general a requirement of semantic nominalism; and indeed in the case of second-order abstraction the representatives are of a lower type altogether than the entities whose equivalence classes they represent. Similarly one might insist that (were we actually to embrace the framework of trans-world abstraction) representatives need not exist at the *same world* as the lines whose classes they represent (the direction of line a at world w might be identified with an object b—perhaps itself a line, perhaps not—at world w').

Thus, we are left with a viable construal of logicism, given by the combination of semantic nominalism and a naturalistic conception of abstraction. It is a form of logicism because cardinality notions, as embodied for instance in the Frege quantifier characteristic of the theory AFQ can be reasonably construed as logical notions. However, this kind of logicism is *non-reductive* in character: no reduction to Hume's Principle is needed in order to establish the logical nature of the notion of cardinality. On the contrary, Hume's Principle is regarded as properly mathematical, and not logical. At the same time we avoid the temptation proffered by the abstraction mystique, and we show how, at least in the case of arithmetic as reconstructed in AFQ (but hopefully more in general in the case of any theory whose abstract entities are delivered by an abstraction principle), one can develop an empirically acceptable theory of mathematical entities. Unsurprisingly, the mathematical development is independent of the philosophical construal of the abstraction principles and the ultimate nature of the objects so introduced.

References

Antonelli, G. A. (2010a). The nature and purpose of numbers. *Journal of Philosophy*, *107*(4), 191–212.
Antonelli, G. A. (2010b). Notions of invariance for abstraction principles. *Philosophia Mathematica*, *18*(3), 276–292.
Antonelli, G. A. (2010c). Numerical abstraction via the Frege quantifier. *Notre Dame Journal of Formal Logic*, *51*(2), 161–179.
Antonelli, G. A. (2012). A note on induction, abstraction, and Dedekind-finiteness. *Notre Dame Journal of Formal Logic*, *53*(2), 187–192.
Antonelli, G. A. (2013). On the general interpretation of first-order quantifiers. *Review of Symbolic Logic*, *6*(4), 637–658.
Antonelli, G. A. (2015). Life on the range: Quine's thesis and semantic indeterminacy. In A. Torza (Ed.), *Quantifiers, quantifiers, and quantifiers: Themes in logic, metaphysics, and language, synthese library 373* (pp. 171–189). Chan: Springer.
Antonelli, G. A., & May, R. (2012). A note on order. In T. Graf, D. Paperno, A. Szabolcsi, & J. Tellings (Eds.), *Theories of everything: in honor of Ed Keenan*, UCLA Working Papers in Linguistics. (Vol. 17, pp. 1–6).
Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*, *70*(19), 661–679.
Bonnay, D. (2008). Logicality and invariance. *Bullettin of Symbolic Logic*, *14*(1), 29–68.
Burgess, J., & Rosen, G. (2005). Nominalism reconsidered. In S. Shapiro (Ed.), *The Oxford handbook of philosophy of mathematics and logic* (Chap. 16, pp. 515–535) Oxford: Oxford University Press.
Butts, R., & Hintikka, J. (Ed.). (1977). *Logic, foundations of mathematics, and computability theory*. Dordrecht: Reidel.
Colyvan, M. (2001). *The indispensability of mathematics*. Oxford: Oxford University Press.
Dedekind, R. (1988). *Was sind und was sollen die Zahlen*, Vieweg: Braunschweig, 1888, English translation, (The Nature and Meaning of Numbers. In R. Dedekind, *Essays on the theory of numbers*, Dover: New York. pp. 29–115.)
Dummett, M. (1991). *Frege: philosophy of mathematics*. Massachusetts: Harvard University Press.
Feferman, S. (1977). Categorical foundations and foundations of category theory. In R. E. Butts & J. Hintikka (Eds.), *Logic, foundations of mathematics, and computability theory, The University of Western Ontario series in philosophy of science* (Vol. 9, pp. 149–169). Springer.
Field, H. (1980). *Science without numbers*. Princeton: Princeton University Press.
Field, H. (1989). *Realism, mathematics and modality*. Oxford: Basil Blackwell.
Florio, S., & Linnebo, Ø. (forthcoming). On the innocence and determinacy of plural quantification. *Noûs* doi:10.1111/nous.12091
Frege, G. (1980). *The foundations of arithmetic: A logico-mathematical enquiry into the concept of number*. (J. L. Austin, Trans.). Evanston: Northwestern University Press.
Goodman, N., & Quine, W. V. (1947). Steps toward a constructive nominalism. *Journal of Symbolic Logic*, *12*(4), 105–122.
Hale, B., & Wright, C. (2001). To bury Caesar. In B. Hale & C. Wright, *The reason's proper study* (pp. 355–396). Oxford: Oxford University Press.
Hale, B. (2001). Is platonism epistemologically bankrupt? In B. Hale & C. Wright, *The reason's proper study* (pp. 169–188). Oxford: Clarendon Press. Originally in *The Philosophical Review*, *103*(2), 299–325 (1994).
Hellman, G. (1989). *Mathematics without numbers: towards a modal-structural interpretation*. Oxford: Clarendon Press.
Higginbotham, J. (1998). On higher-order logic and natural language. In T. Smiley (Ed.), *Philosophical Logic, Proceedings of the British Academy* (Vol. 95, pp. 1–27).
Hilary, P. (1971). *Philosophy of Logic*. New York: Harper essays in philosophy, Harper & Row.
Weierstraß, K. (1854). Zur Theorie Der Abelschen Functionen. *Journal für die reine und angewandte Mathematik*, *47*, 289–306.

Lewis, D. (1991). *Parts of classes.* Hoboken: Blackwell.
Panza, M., & Sereni, A. (2013). *Plato's problem: An introduction to mathematical platonism*. Basingstoke: Palgrave Macmillan.
Peters, S., & Westerståhl, D. (2006). *Quantifiers in language and logic*. Oxford: Clarendon Press.
Prior, A. (1971). P. T. Geach & A. J. P. Kenny (Eds.), *Objects of thought*. Oxford: Clarendon Press.
Quine, W. V. (1951). Ontology and ideology. *Philosophical Studies*, *2*(1), 11–15.
Rosen, G. (1993). The refutation of nominalism (?). *Philosophical Topics*, *21*(2), 141–186.
Russell, B. (1918–1919). The philosophy of logical atomism. *The Monist*, 28 and 29: 495–527 (part I) and 33–63 (part II).
Tarski, A. (1986). What are logical notions?. *History and Philosophy of Logic*, *7*(2), 143–154.
Wright, C. (1983). *Frege's conception of numbers as objects*. Aberdeen: Aberdeen University Press.

Chapter 3
Discussion Note On: “Semantic Nominalism: How I Learned to Stop Worrying and Love Universals” by G. Aldo Antonelli

Robert C. May and Marco Panza

Editorial Note The following Discussion Note is an edited transcription of the discussion on G. Aldo Antonelli’s paper “Semantic Nominalism: How I Learned to Stop Worrying and Love Universals” (this volume), held among participants at the IHPST-UC Davis Workshop *Ontological Commitment in Mathematics* which took place, *in memoriam* of Aldo Antonelli, at IHPST in Paris on December, 14–15, 2015. The note’s and volume’s editors would like to thank all participants in the discussion for their contributions, and Alberto Naibo, Michael Wright and the personnel at IHPST for their technical support.

Brice Halimi

I have three remarks.

The first one is just about the first page of the paper and the quotation of Lewis. I think that it catches up with what Andy Arana said. In a way, the logician cannot say what is impossible, and so the philosopher cannot rule what mathematics has to do. But the philosopher and the logician can at least say what is possible. And I think that it is dangerous to separate too sharply what philosophy rules and what mathematics does. Think of Wittgenstein and of the conception that he had of a two-way connection between philosophy and mathematics.

The second remark has to do with the meta-theoretic framework involved in Aldo’s construction. It looks like it is simple type theory and set theory at the same time. My problem is that Henkin’s results rely on simple type theory, while Tarski’s paper on logical notions relies on set theory. So the perplexity I have concerns the framework used in this paper. I have no clear answer, but it reminds of a paper by Philippe de

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

M. Panza
CRNS, IHPST (CNRS and University of Paris 1, Pantheon-Sorbonne),
Presidential Fellow at Chapman University, Paris, France
e-mail: marco.panza@univ-paris1.fr

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_3

Rouilhan—who is a very well-know scholar, here in France—that appeared in the *Proceedings of the Aristotelian Society* (De Rouilhan 2002). He claims that plural quantification should range over pluralities. The reason is that, otherwise, set theory would claim everything, it would capture everything, and everything would end up being first-order because in set theory everything is first-order. So that's my problem: either you endorse simple type theory and then universals are smuggled in, because of course in your meta-theoretic framework you have universals; or you endorse set theory but then everything is first-order, which is trivializing in a way.

My last remark is about the mappings that are referred to in the paper, in particular on page 10 and 17. They assign a representative to each universal within the first-order domain. I think that the ontological and the epistemological status of these mappings should be questioned. In his paper Aldo reminds us of Quine's distinction between ontology and ideology. I think we should add a third item, maybe technology? A mapping is not something that is part of ontology properly speaking, it is not part of ideology, but those mappings which are required to make sense of our use of universals are something, both epistemically speaking and ontologically speaking. And they are parts of the apparatus, the global apparatus. I think it is something that should be accounted for.

Robert May

Let me make a remark on Brice's first point about the quotation from David Lewis. It seems to me that Lewis' remark reflects a peculiarly of the modern, contemporary perspective. If you go back and look at late 19th century mathematics, especially in Germany, you see that there are conceptual ideas that are pressing mathematicians. This is certainly so for Frege, but also for Dedekind and Hilbert, for example. I think this was indicative of that time, but you find something similar in other places in the history of science, where the practitioners of the science are driven by conceptual foundations, by questions that are genuinely philosophical. The tenor now is to divide the responsibility, so mathematicians and scientists do not typically think that it is part of what they do to start from conceptual foundations and see how mathematical and theoretical structure reflects or embodies the underlying conceptual view. But when these conceptual ideas are detached from the actual workings of the inquiry, they become purely philosophical, and take on a life of their own. So, I think that Lewis's quote reflects that change in attitude that appears when you study history of science, which is punctuated by periods in which great advances are made precisely because the actual work is flowing from an underlying set of conceptual worries. Frege was concerned with what is a number. To this end he put in conceptual foundations for what he was doing. And he was not alone. This was a style of thinking which was common in his day, in mathematical and logical and scientific circles.

Fernando Ferreira

I enjoyed very much this paper of Aldo's. I didn't know Aldo very well. I met him for the first time, I think, in Irvine in 2004, and then I met him a couple of times afterwards. When I read his paper and his theory of arithmetic with the Frege quantifier, at first I thought it couldn't work. How can one derive all the first-order Peano–Dedekind axioms from his axioms, I wondered. But then I saw how things worked. And I think it's very interesting because he only derives first-order Peano arithmetic as opposed, for instance, to second-order Peano arithmetic, drawn from Hume's Principle. He does not quantify over concepts, indeed. So, it's a very ingenious system, and I tend to agree with him about the emphasis he makes on ideology as opposed to ontology. It's a very ideological system because he has the Frege quantifier. However, he apparently has to make some concessions to ontology because he still has the number operator that goes from concepts to objects. He also makes some concessions to ideology that are not truly logic, as it is the case with the predicate N.

I also found very interesting his defense of the view that reference of abstract terms is not determinate (I don't remember well, in fact, whether these are the words that he actually uses). It seems that determinate reference of abstracts is kind of instrumental. And this I found very interesting. So these are comments, rather than questions.

Marco Panza

I'm wondering in which sense the position that Aldo defends in this paper is nominalist about reference. It is something that is not so clear to me, and not evident at all. Because, after all, even if we accept that abstraction principles are innocent, in some sense (even if they are not logical but mathematical principles), it remains that in the presence of comprehension they may deliver the existence of objects and in fact, if the equivalence relation on the right-hand side satisfies certain very weak and simple conditions, abstraction principles deliver the existence of countably many objects.

So, to say that this is a nominalist position depends, in the end, on the fact that admitting abstraction principles is, so to say, a nominalistically legitimate way to introduce abstract objects and to lead to the assertion of their existence. Still, this is a point that is difficult for me to understand unless we assert that there is something like a special sense of existence that abstraction principles deliver. But this is not something that it seems to me Aldo was really ready to accept. In any case it is not something that he is arguing for in his paper, and according to my knowledge, in no other of his papers.

But, if there is no special sense of existence, I don't see why, we should say that Aldo's views are nominalist. Sure, the objects that we introduce and of which we prove existence by abstraction principles are just those that we use in order to do arithmetic. And they are not completely determined as Fernando just said. But why not just say that this is, after all a form of Platonism? I don't see exactly the point of defending the idea that this is a nominalist position.

This is something that I would have liked to ask him, but I did not have the time. So I leave this question open for future comments or remarks.

Elaine Landry

I might be able to help out here, since Aldo and I had conversations about this exact point. There are two things to be said.

The first is that the phrase "semantic nominalism" came out of a reaction to what was called "semantic realism", which was just robust realist-platonism. This is something that does not appear in the paper. So someone might have to do some research to look it up. But I think that the phrase comes out in Rosen, and I think the contrast was important for Aldo.

More importantly I think he was pressed to answer the question "why is this not just a notion of existence in a Platonistic sense? What is special about abstraction principles?". Concerning this, I think that the crucial point is that here there is no concession to the idea that we have independent existence. If you are a Platonist, you take abstract objects to exist independently of us and we grasp them, and the way that we grasp them might be abstraction principles. That would make the story Platonist. But that is not the story he is telling. The story he's telling is that abstraction principles are yielding existence claims for you; abstraction principles and existence go hand in hand. By these principles, we are not going to access otherwise existing things. It's the abstraction principle that is yielding, legitimizing, the relevant objects. The objects and the principles go together, as it were.

Robert May

It's as if he almost went over the boundaries into an explicit definition. The way he put abstraction principles, they are almost functioning like explicit definitions, and Aldo is trying to say: no, that's too far, to make them like explicit definitions.

The standard paraphrase of abstraction principles is revealing. We paraphrase Hume's Principle, for instance, as saying that a number is *whatever* equinumerous concepts have in common. That's the kind of sense of indeterminacy that Fernando was gesturing towards. Aldo's idea is this: yes, there is a notion that there is something, but you are not making a claim about what the nature of that thing is, it is whatever meets this condition. So it leaves open precisely the question of whether there are abstract objects *qua* the referents of abstract term. He leaves that question open. And this is a relevant question.

I think the point that Aldo makes towards the end of the paper is that nothing in the nature of the way you answer, or specify the "whatever" clause is going to matter for Frege's theorem. I think that's an important insight. Nothing in the role that Frege's theorem plays in the derivation of Peano axioms depends in any way on the decision you make about what exactly numbers are, so long as they are whatever equinumerous concepts have in common.

Marco Panza

I perfectly understand this point. But according to my reading (but, of course my reading may be wrong), there is absolutely nothing in the classic neologicist position that forces us to another view: the fact that numbers exist independently is simply an ideological point that neologicists can make, but they are no way forced to make because of their mathematical and/or logical achievements.

So, if this is the point, what Aldo is saying is this, it seems to me: look, what the neologicists tell us cannot force us to accept the independent existence of numbers, and so we can read the neologicist project in a weak way, in a completely neutral manner.

I agree. But in order to arrive there, the technical results that Aldo presents in this paper—or better that he presents in other papers and mentions in this—are not necessary. They are certainly very interesting results, but it seems to me that they are not needed at all in order to make this point. It can be simply made by looking at Hale and Wright’s construction and saying: What you say about existence does not follow from your technical result; I do not know whether it is true or false, right or wrong. Simply it does not follow.

Robert May

Neo-logicists, I think, want to press the point because of the epistemic role they have in mind for abstraction principles, and in part what Aldo is trying to step away from is the commitment they have to a certain epistemic role for Hume’s Principle.

Giorgio Venturi

There is probably a higher challenge for Aldo’s views to be really nominalist. The first time I read this paper my idea was this: Here is a statement that makes perfectly good sense in a structuralist context. This is shown by passages like:

> On the naturalistic view of abstraction, numbers are whatever entities are selected as representatives of equivalence classes of concepts.

Or even:

> Our knowledge of […][abstract entities] is limited, but still substantial. In particular, we know whatever has to be true no matter how the representatives are chosen.

The particular nature of objects does not really matter, but still the structure is there in some sense.

So the point is perhaps where to put the accent on existence when we come to semantics (and for me it was very instructive to hear from Elaine that Aldo’s position is a reaction to semantic realism). What I mean is that the real question concerns the type of reference we have in our semantics. And Marco’s question is actually relevant to this point, because the point is what is the background theory that we have, since if we just assume set theory then this is real structuralism. But it is not, of course.

Now, as far as I can understand, what Aldo is proposing is that there is probably a more linguistic reading of semantics in this sense. In the sense that the justification of our universals is in the end given by our use of them in our language, by our proficiency in using them in mathematical or other contexts. So to make Aldo’s proposal even stronger, one should clarify what is the nature of the semantic notions, and what is reference as a semantic notion, not as an ontological one. The question at issue is then the difference between ontology and semantics. In some accounts that just collapses, as it happens, for instance, in the model-theoretic account.

Paolo Natali

Probably this is simple-minded, but my impression is that this paper somehow deals more with the metaphysics of abstracts rather than with their ontology. So in a sense it is ontologically realist but it objects to some particular views of what metaphysically corresponds to that.

I am just using the distinction between ontology and metaphysics as a distinction between disciplines that inquire, respectively, about what there is and about how things are. My impression is that Aldo's position is that of an ontological realist that wants to object to a substantive view of the metaphysics of abstracts; that is, of someone who is concerned with denying that *abstracta* have too loaded a metaphysical status. So I think that Marco's point about Aldo not having a nominalist account of existence, in the sense of an ontologically nominalist position, stands as correct.

Andrea Sereni

I'd have several comments, but I will confine myself to a couple of them and then I'll make a point that is connected to what Marco and partially Giorgio said.

First, I think one of the crucial aspects in the dialectic that Aldo wants to set up with the neologicists is the following: neologicists would press the point that mathematical evidence is not enough in a foundational project, and that definitions, especially Hume's Principle, should be conceived as principles for the individuation of mathematical objects, aimed at characterizing in a definite and univocal way their individual nature. And of course, this is something that Aldo is calling into question. The dialectic here seems analogous to what Benacerraf (and this is connected with Giorgio's point) suggested in his 1965 paper (Benacerraf 1965) and with Wright's reply to it in his 1983 book (Wright 1983): the issue is whether the evidence that we shall use in explaining mathematical knowledge should come from mathematics alone or should also be informed by other metaphysical or epistemic constraints. This seems to me an important part of what distinguishes Aldo's position from the Neo-Fregean.

Now, coming to my second comment, Marco asked whether we should really conceive of Aldo's position as nominalistic. I think there is a way of answering this question positively by considering a very simple thought experiment. In a Neo-Fregean picture if there were, *per impossible,* no numbers, no mathematical objects, then we wouldn't have Frege's Theorem. In Aldo's picture, on the contrary, if we didn't have abstract mathematical objects we may still have all the mathematics we have, since we can still make use of concrete representatives for those objects.

Marco Panza

But then these concrete representative have to exist.

Andrea Sereni

Yes, this goes exactly in the direction of the point I'd like to raise.

There is something Aldo says at a certain point that I found very interesting, especially as it seems to me to clash with the naturalist viewpoint that he wants to uphold. He says that in order for his position to work, Hume's Principle should be taken as a way of replacing an axiom of infinity. This claim stands in clear contrast to what happens in the neo-logicist picture; Neo-logicists allow for situations in which we have a finite concrete domain and an infinite abstract domain. On the contrary, it seems to me that in Aldo's picture one is bound to use Hume's Principle to provide an infinite domain with infinitely many concrete objects working as representatives for the natural numbers. But then, one concern the naturalist can have is the following: Hume's Principle can hardly be a naturalistically acceptable means of introducing an infinity of concrete objects just because there are no naturalistically acceptable means of introducing an infinity of concrete objects at all—apart from, maybe, empirical findings coming from physics.

Fernando Ferreira

I tend to agree with Marco that Aldo's position is not nominalistic, but it is also not crude Platonism. As Paolo said before, it is about the metaphysics of the abstract objects. So, while Platonism is like postulation, abstractionism is not like that, because objects come out from abstraction, and hence operate by means of the number operator and certain kinds of ideological notions, like the Frege quantifier. And, insofar as they follow from these notions, somehow they provide an answer to the epistemological question. Mere Platonism does not seem to be able to give such an answer.

On the other side, given the infinitude of numbers, an abstractionist cannot be a nominalist unless you say that the universe contains infinitely many objects. Once you've been given the objects, you just say that numbers correspond to any of these objects, no matter which ones. The existence of infinitely many objects is a contingent thing. Nevertheless, Aldo's view is no mere crude Platonism, I think.

One can see the idea of Neo-logicism, when one compares first-order abstraction with second-order abstraction. In first-order abstraction you actually do not need the objects (I think), you can reason in context and rewrite everything without the need of the abstractions. It's only when you go to second-order abstraction that you really are getting something new in terms of ontology.

Alberto Naibo

As a proof-theorist, I found Sect. 4 of Aldo's paper very interesting, and in this context I would like to come back to some points made by Giorgio.

The case Aldo makes for our understanding of generalized quantifiers is somehow linked to our human inferential abilities or capacities, and, indeed, he mentions two examples of inference suggesting it. It seems to me there is a way of interpreting semantics as being concerned with use, in a certain sense. And this is also connected to Aldo's quoting Feferman's paper, because Feferman is there speaking of operations, that is, I'd like to say, of operative semantics.

Maybe a way to try to understand whether there is some kind of operative semantics behind Aldo's paper is to look at what he says about a closure condition on a domain D_2 linked to the permutations on a domain D_1. Since there is perhaps a connection to make between this invariance criterion for logicality and the uniqueness condition that we have, for example, in inferentialist semantics, namely Belnap's uniqueness; that is, really a link between an invariance condition and a uniqueness condition in the sense that what is required is the unicity of a certain operator, and this is maybe something that makes such an operator logical, as opposed to mathematical operators.

Francesca Boccuni

Just a quick follow up to what Giorgio was talking about before, especially when he wondered what kind of a notion of reference is underlying Aldo's view.

I think that (and this is also connected with the opposition between metaphysics and ontology, or between essence and existence, so to say) the most natural understanding—to me—of the way the notion of reference works in Aldo's paper is that reference is just a function. It is heavily loaded metaphysically neither with the salience of reference itself, nor with the salience of the object that a name purportedly refers to. Basically, what one needs to have reference for is mapping a language to a non-linguistic domain.

This very notion does not necessarily go together with the need to individuate, in a metaphysically committal way, the objects one wants to refer to and that's pretty much it. In philosophy of language, there is a long-going debate on reference having salient features tightly connected with, e.g., intentionality, individuation, existence and uniqueness. But, in fact, all one needs is just a function. So, even if set theory provides the semantics for that kind of formal framework, from that one should only conclude that set theory just provides *a* model for reference. But different models might be provided. So even if one chooses set theory in order to provide a model, this does not say anything special about the nature of reference. It's just a mapping.

Giorgio Venturi

Just a quick comment on this notion of reference. The point is that when you buy set theory, then you have to specify what a function is. One can surely say that reference is just a function, but, then, either one gives a set-theoretic account of functions, according to which a function is just a set of pairs, or one needs, as it were, a larger notion of a function, a notion which is not set-theoretical and fits with the idea that reference depends on our ability of using language. But then the problem is that, if one takes the latter option, it becomes very hard to build a logicist approach to mathematics, because there is a clear element of psychologism in such an option, and one has to account for this, or, at least, give an argument for the view that this is our cognitive way of doing things. But, then, either one buys psychologism or Kantism, or a form of both.

Robert May

I have a remark inspired by Francesca's. She made an interesting contrast between robust and narrow notions of reference. Whenever we look at natural language, we immediately go to all the things she wants to eliminate, all the things about intentionality, etc. that are tied to a robust notion of reference. On the other hand, when we look at mathematics and logic, we fall into the narrow notion. But that's interesting when we look at quantification, because the notion of quantification holds across the board. The notions of quantification that we need and utilize within the logico-mathematical setting are also perfectly appropriate within a natural setting.

For example, generalized quantifiers are taken by linguists as being an appropriate and adequate theory of natural language semantics.

In Aldo's paper, all the examples from natural language have to do with quantifiers, because it is the way we look at the notion of quantification that really seems to run the gamut across all languages. The notion of reference really changes, and when look into logic and mathematical domains, we really want the narrow, lean notion; we want to abstract away from the things that Francesca was saying. But this really leads to something very different.

I think something like this is part of what Russell was worried about in "On Denoting" (Russell 1905), when he talks about the Grey's Elegy argument. There he is trying to argue that there is something funny about the way Frege utilizes the notion of reference. For Russell, reference isn't really a legitimate notion within logic and mathematics, and so we need a theory of descriptions which is motivated within the logical system relative to the notion of generalization.

So I think the intuition about reference is somehow funny when we compare the notions of reference we utilise in logical and mathematical contexts, with that which we draw out from our understanding of natural language. In contrast, there is no oddity with respect to quantification, which as noted seems to run the gamut from natural to logical languages (and again this is Russell's understanding as well).

References

Benacerraf, P. (1965). What numbers could not be. *The Philosophical Review*, *74*, 47–73.

De Rouilhan, P. (2002). On what there are. *Proceedings of the Aristotelian Society*, *102*, 183–200.

Russell, B. (1905). On denoting. *Mind*, *14*, 479–493.

Wright, C. (1983). *Frege's conception of numbers as objects*. Aberdeen: Aberdeen University Press.

Chapter 4
Semantic Assumptions in the Philosophy of Mathematics

Robert Knowles

Abstract The standard semantic analysis of sentences such as 'The number of planets in the solar system is eight' is that they are identity statements that identify certain mathematical objects, namely numbers. The analysis thereby facilitates arguments for a controversial philosophical position, namely realism about mathematical objects. Accordingly, whether or not this analysis is accurate should concern philosophers greatly. Recently, several authors have offered rival analyses of sentences such as these. In this paper, I will consider a wide range of linguistic evidence and show that all of these analyses, including the standard analysis, suffer significant drawbacks. I will then outline and present further evidence in favour of my own analysis, developed elsewhere, according to which such sentences are identity statements that identify certain kinds of facts. I also defend a novel and plausible approach to the semantics of interrogative clauses that corroborates my analysis. Finally, I discuss how realists about mathematical objects should proceed in light of the arguments presented in this paper.

Keywords The number of planets · Philosophy of mathematics · Semantics

4.1 Truth and Indispensability

The following sentences are true:

(1a) The number of planets in the solar system is eight.
(1b) The mass of Jupiter in kilograms is 1.8986×10^{27}.

This observation appears harmless enough, but such observations facilitate important and controversial arguments in the philosophy of mathematics. For example, if we adopt the standard assumption that the mathematical expressions in (1a–b) stand for mathematical objects, then in order for (1a–b) to be true, those mathematical

---

R. Knowles (✉)
University of Manchester, Manchester, England
e-mail: rfknowles@gmail.com; robert.knowles@manchester.ac.uk

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_4

objects must exist. From this, we get a so-called 'easy argument' for realism about mathematical objects.

Another observation concerning sentences such as (1a–b) is that they are indispensable to our best scientific theories. In spite of Field's (1980, 1989) heroic attempts to show otherwise, we cannot do science without employing arithmetical language. Add to this the premise that we should believe in all those entities apparent reference to which is indispensable to our best scientific theories, and we have the foundations of the indispensability argument for realism about mathematical objects.

Both of these arguments rely on the substantial semantic thesis that sentences such as (1a–b) contain expressions that refer to numbers. The standard analysis of (1a–b) assumed in much of the philosophy of mathematics says that number-of expressions and magnitude-of expressions, such as 'The number of planets in the solar system' and 'The mass of Jupiter in kilograms', as well as numerals, such as 'eight' and '1.8986×10^{27}', refer to numbers. On this view, sentences such as (1a–b) are naturally interpreted as expressing the identity of the referents of the pre- and post-copular expressions. Because it is inspired by the work of Frege (1884) 1953, p. 69, I will call this analysis the 'Fregean analysis'. I should distinguish here between the Fregean analysis itself and the spirit in which it can be put forward. This is important, since Frege arguably did not develop the analysis in order to account for the semantics of natural language, but instead as part of his project of developing an ideal language for mathematics and science. In this paper, it is the Fregean analysis as put forward as a thesis about natural language that concerns me.

It is easy to see how the Fregean analysis facilitates easy arguments. If (1a) states that the number of planets is identical with the number eight, then its truth-conditions demand that the number eight exists. Hence, from the plausible premise that (1a) is true, we reach the conclusion that there are numbers.

The way in which indispensability arguments rely on the Fregean analysis requires further explanation. The most influential, and arguably the first, presentation of the indispensability argument is in Quine (1948). For Quine, to establish what a theory says there is we must first translate the theory into a canonical language. For familiar reasons, first-order predicate logic is chosen for this task. Then we establish which objects the bound variables of the theory must range over in order for the theory to be true. Those objects are what the theory is committed to. Indispensability arguments rest on the principle that the entities that our best scientific theories are committed to in this way exist.

For the indispensability argument presented above to go through, it must be assumed that the indispensability of sentences such as (1a–b) to our best scientific theories implies that numbers must be among the values of the bound variables in our canonical formulations of these sentences in order for the theories they comprise to be true. This assumption is borne out by the procedure used to translate our scientific theories into the canonical language, and, in particular, how we translate referring expressions. Quine does not propose that we introduce a constant into the language for each referring expression, but instead shows how to capture the truth-conditions of sentences containing referring expressions with just predicates, variables and first-order quantifiers. To take the famous example, 'Pegasus flies' is not translated into

'Fp', where 'p' is a constant that refers to Pegasus and 'F' is the predicate '…flies'. Instead, it is translated as '$\exists x(Px \& \forall y(Py \rightarrow x = y) \& Fx)$', where 'P' is the predicate '…pegasizes'. For a theory containing this sentence to be true, Pegasus must be among the values for the bound variables, and so a theory containing this sentence is committed to the existence of Pegasus.

The above indispensability argument therefore assumes that sentences such as (1a–b) contain expressions that refer to numbers. Without this assumption, there is no reason to think that, for the first-order counterparts of (1a–b) to be true, a number must serve as the value of the bound variables.

We have seen that the Fregean analysis facilitates arguments for a controversial philosophical position. Accordingly, whether or not this analysis is accurate should concern philosophers greatly. Recently, several authors (Hofweber 2005; Moltmann 2013a, b, forthcoming; and Felka 2014) have offered rival analyses of sentences such as (1a–b). In what follows, I will consider a wide range of linguistic evidence demonstrating that all of these analyses suffer significant drawbacks. I will then outline and present evidence in favour of an alternative analysis.

In Sect. 4.2, I present a syntactico-semantic puzzle, *Frege's Other Puzzle*, that any plausible analysis of (1a–b) must provide a solution to, and show that the Fregean analysis fails to do so. I endorse an existing attempt to solve the semantic dimension of this puzzle, which involves arguing that the post-copular expressions of (1a–b) are not referring expressions. I show that evidence provided by Friederike Moltmann in favour of this claim is inconclusive, but provide my own evidence that is more suggestive. In Sect. 4.3, I discuss evidence presented by Moltmann that is supposed to show two things: first, that the pre-copular expressions in (1a–b) do not purport to refer to numbers; second, that they do purport to refer to tropes. I show that her evidence is only suggestive of the former conclusion and provide some evidence to suggest that number-of and magnitude-of expressions instead refer to facts. In Sect. 4.4, I present a recent syntactic analysis of sentences such as (1a–b) as *specificational sentences* comprised of question and answer pairs that promises to solve the syntactic dimension of Frege's Other Puzzle. Though I agree that (1a–b) are specificational, I raise some issues with the question-answer analysis. In Sect. 4.5, I argue for a particular analysis of questions and answers that corroborates two claims of the preceding sections: that the received semantic analysis of specificational sentences is wrong; and that the pre-copular expressions in (1a–b) refer to facts. In the final Sect. 4.6, I present a satisfactory semantic analysis of specificational sentences, and so of (1a–b). I show that philosophers cannot appeal to the truth-conditions of sentences such as (1a–b) alone in order to establish realism about mathematical objects. I end by suggesting ways in which realists might proceed in light of this: they must either supplement their arguments with metaphysical arguments, or appeal to different areas of mathematical language. I should note that, in what follows, I will use '?' to indicate that a sentence is ungrammatical, and '*' to indicate that a sentence is semantically deviant, or otherwise infelicitous.

4.2 Frege's Other Puzzle and the Semantics of Numerals

Hofweber (2005, pp. 179–180) identifies a puzzle for the Fregean analysis. Numerals can occupy two syntactic positions in natural language sentences. In (1a), the cardinal 'eight' is in singular-term position, paradigmatically occupied by names, such as 'Dostoyevsky':

(2) The author of Crime and Punishment is Dostoyevsky.

Contrast (3a–b), in which the mathematical expressions are in adjectival position, typically reserved for adjectives, such as 'red':

(3a) There are eight planets in the solar system.
(3b) Jupiter has a mass of 1.8986×10^{27} kg.
(3c) There are red planets in the solar system.

The puzzle has a syntactic and semantic dimension. The syntactic dimension is that grammar forbids referring expressions from occurring in adjectival position. For example, replacing 'eight', '1.8986×10^{27}', and 'red' in (3a–c) with 'The number of planets in the solar system', 'The mass of Jupiter in kilograms', and 'The colour of blood', respectively, yields ungrammatical results:

(4a) ? There are the number of planets in the solar system planets. (Adapted from Hofweber (2005, p. 179))
(4b) ? Jupiter has a mass of the mass of Jupiter kilograms.
(4c) ? There are the colour of blood planets in the solar system.

Intuitively, 'The number of planets in the solar system', 'The mass Jupiter in kilograms' and 'The colour of blood' cannot take adjectival position because they are referring expressions, making it all the more puzzling that numerals can.

The semantic dimension of the puzzle arises because numerals in adjectival position also appear to have a different semantic function to their counterparts in singular-term position: they do not appear to stand for an object. An account of the relation between the semantic function of both occurrences must be provided. Following Hofweber (2005, p. 180), I call this syntactico-semantic puzzle 'Frege's Other Puzzle' (FOP) (again, given Frege's project, I do not wish to imply that Frege himself was or should have been concerned with this puzzle; this is a puzzle for the Fregean analysis put forward as a thesis about natural language semantics).

Hofweber (2005), Moltmann (2013a, b) and Felka (2014) have motivated analyses of sentences such as (1a–b) by claiming that they solve FOP. A semantic theory that solves FOP is to be preferred to the Fregean analysis only if it provides the best explanation of all the other available linguistic data. We shall see that none of these analyses do both.

Though I will reject their analyses, I endorse the spirit of these authors' solutions to the semantic dimension of FOP, which involves arguing that the post-copular expressions in (1a–b) are not referring expressions. Generalised quantifier theory (GQT) yields a plausible interpretation of numerals in adjectival position as determiners. (GQT is based on the work of Mostowski (1957), Montague (1974), and has subsequently been developed by many theorists. See Keenen and Westerståhl (1997) for an overview.) Along with a demonstration that numerals function as determiners in contexts such as (1a–b), this is sufficient to solve the semantic dimension of FOP. (However, see Knowles (2015, pp. 2758–2762) for a GQT-based account of numerals as adjectives. The overall argument of this paper does not rest on which specific semantic function is assigned to the numerals in (1a–b), but only that they do not serve as referring expressions.)

In the remainder of this section, I will do three things. First, I will outline the GQT account of numerals. Second, I will show that evidence presented by Moltmann in support of the claim that the numerals occurring in (1a–b) are not referring expressions is inconclusive. Third, I will present my own evidence for this claim that is more suggestive.

Analysing natural language using the quantifiers of predicate calculus is problematic. 'Something' is assigned the existential quantifier and 'Everything' the universal; these are the basic units of analysis. But in English there are more complex quantifier phrases, such as the following: 'Some apples'; 'Every man'; 'Few coins' etc. This suggests that 'Something' and 'Everything' are also complex, made up of 'Some', 'Every' and 'Thing' (I am grateful to an anonymous reviewer for pointing out that the universal quantifier in other languages, such as French and German, does not behave in this way). There are also quantifier phrases in natural language that resist analysis in terms of the existential and universal quantifiers ('Most', for example). An adequate analysis of natural language should provide a unified and systematic analysis of quantifier phrases (cf. Hofweber 2005, p. 196), and account for all the quantifier phrases of the language (cf. Barwise and Cooper 1981, pp. 159–161).

GQT provides such an account. According to GQT, sentences are typically composed of two syntactic types: a verb phrase (VP) and a noun phrase (NP). Sets are the semantic values of VPs and functions from sets to truth-values are the semantic values of NPs. The set of things that run fast is assigned to 'run fast'; assigned to 'Adult cheetahs' is a function that yields true for all and only sets that contain the adult cheetahs. Thus, 'Adult cheetahs run fast' is true iff the adult cheetahs are members of the set of things that run fast; that is, true iff adult cheetahs run fast.

NPs can be complex, consisting of a noun and a modifier. Determiners modify nouns or NPs to help determine the extent of their reference. 'Some' in 'Some apples' modifies 'apples' and determines reference to at least one apple. Other determiners include 'Most', 'Every' and 'Both'. Quantified NPs are complex NPs whose modifying element is a determiner. The semantic value of a noun is a set so the semantic value of a determiner is a function from sets to functions from sets to truth-values.

According to GQT, numerals in adjectival position are determiners. They take NPs as arguments and determine their reference: 'Six' determines reference to six apples in 'Six apples'. This is a plausible and widely accepted semantics for determiners that yields a plausible semantic account of numerals in adjectival position.

Recall (1a–b) and (3a–b):

(1a) The number of planets in the solar system is eight.
(1b) The mass of Jupiter in kilograms is 1.8986×10^{27}.

(3a) There are eight planets in the solar system.
(3b) Jupiter has a mass of 1.8986×10^{27} kg.

Either 'eight' and '1.8986×10^{27}' must be shown to have related semantic functions across these two contexts, or the difference must be plausibly and systematically explained. After all, the numerals in (1a–b) and the numerals in (3a–b) are not plausibly homonyms. To meet this challenge, Moltmann (2013a, p. 522) provides evidence for the claim that sentences such as (1a–b) are not identity statements, thus implying that their post-copular expressions are not referring expressions:

(5) *The number of planets in the solar system is the number eight.

If 'The number eight' and 'eight' are both singular terms standing for the number eight, they should be substitutable *salva congruitate* and *salva veritate*. However, (5) sounds odd, and the truth of (1a) is not obviously preserved in it. The supposed explanation of these intuitions is that the two expressions belong to different semantic types. If 'eight' in (5) is a determiner, for example, then the unacceptability of the substitution would be expected.

This evidence is inconclusive for two reasons. First, there is a simpler explanation: (5) is a pleonastic sentence. The repetition of information in the latter part could suggest it is just a long-winded version of (1a). Compare:

(6a) The metal of the bike is steel.
(6b) The metal of the bike is the metal steel.

(6b) is odd because it repeats information, and is rarely, if ever, used. It does not follow that 'steel' is of a different semantic type to 'the metal steel'; quite the opposite: the former is the latter abbreviated. This can explain why (5) sounds odd to us, and its oddness, in turn, can explain why it is not obvious that the truth of (1a) is preserved.

Second, the principle that Moltmann's evidence relies on, that co-referring expressions must be directly substitutable *salva congruitate* (call it the 'naïve reference principle') has been discredited (see Oliver 2005, among others). Co-referring expressions are often not substitutable, so (5) does no more than further demonstrate the inadequacy of the naïve reference principle.

However, Dolby (2009) convincingly defends a modified version of the principle: co-referring expressions are substitutable by a process of generalisation and specification in accordance with the rules of grammar. Call this the 'reference principle'. I will now defend this principle, and use it to provide more compelling evidence that suggests numerals do not behave as referring expressions in contexts such as (1a–b), and so lends support to the analysis of them as determiners.

Dolby helpfully categorises failures of direct substitution into three categories: first, those involving complex referring expressions; second, those problematic because of the inflection of language; and third, those involving pre-modifying adjectives. I will concentrate on the third, most relevant kind.

To use Dolby's example, 'Multi-cultural' occurs as a pre-modifying adjective in (7a):

(7a) Multi-cultural Britain has draconian restrictions on free speech.
(7b) ? Multi-cultural the land of Churchill has draconian restrictions on free speech.

Replacing the co-referring 'The land of Churchill' for 'Britain' gives the ungrammatical (7b).

Dolby claims that co-referring expressions often cannot be directly substituted because the rules governing substitution are more complicated than the naïve reference principle implies: 'the rules according to which competent speakers make substitutions are also the rules for the formation of generalizations from particular statements and for the specification of these generalizations' (Dolby 2009, p. 90).

Dolby's rules of generalisation and specification are as follows. The singular term is replaced with a general term (a determiner) in accordance with the rules of grammar:

(8a) Hungry James came home and ate some pasta.
(8b) Someone hungry came home and ate some pasta.

Then the generalisation is specified using a co-referring expression for 'James':

(8c) The hungry student of the house came home and ate some pasta.

This is grammatical, but Dolby anticipates an objection:

> If substitution is to be understood as proceeding according to rules that form grammatical sentences from grammatical sentences, how can any substitution ever fail…?[1]

Substitution can fail when there are no rules that take us from the sentence to a generalisation and back to a specification including the expression substituted. For instance, there are no rules for substituting 'Silently' for 'the leader of the free world' in any sentence (Dolby 2009, pp. 294–295).

(5) alone does not support the claim that numerals are not referring expressions in contexts such as (1a–b): there are numerous examples of obviously co-referring expressions not substituting directly. However, we have seen that the naïve reference principle does not reflect the rules for substitution in natural language. Substitution failure occurs when there are no rules taking us from an appropriate generalisation to an appropriate specification. For the substitution of 'The number eight' for 'eight' in (1a) to fail, there must be no rules taking us from a generalisation of (1a) to a specification that includes 'The number eight'. I will now argue that there are no such rules.

There are two ways in which we might generalise from (1a–b). First:

(1a') The number of planets in the solar system is *something*.
(1b') The mass of Jupiter in kilograms is *something*.

To my ear, both sound strange, but even if they are not, we should not be hasty in drawing any conclusions from this. Though 'something' is often apt for being replaced by a referring expression, in some contexts it is not. For example, if Mary and John are virtuous, we can infer:

[1] Dolby (2009, p. 294).

(9a) There is *something* that both Mary and John are.

Specifying what Mary and John are by extending the sentence cannot be done using the referring expression ‘Virtue’:

(9b) ? There is something that both Mary and John are, namely Virtue.

With the predicate ‘virtuous’, however, it can:

(9c) There is something that both Mary and John are, namely virtuous.

This reveals that ‘something’ is not standing in for a referring expression in (9a) (for more on this ‘non-nominal quantification’, see Prior (1971, pp. 34–37) and Sellars (1960); see also Künne (2006, pp. 272–279) and Rosefeldt (2008)).

So, we cannot infer anything from the fact that (1a’) and (1b’) are appropriate generalisations of (1a–b) (if they are) because it is unclear what position ‘something’ occupies. In contrast, the following equally acceptable generalisations are suggestive that it does not stand in for a singular term:

(1a”) The number of planets in the solar system is *so many/however many/that many*.
(1b”) The mass of the rock in kilograms is *so many/however many/that many*.

Some of these sentences may not strike the reader as particularly idiomatic, but in this context, the fact that they are grammatical and not semantically deviant is enough to be suggestive (I am grateful to an anonymous reviewer for pressing me on this point, and for suggesting ‘that many’). ‘So many’, ‘however many’ and ‘that many’ are not plausibly expressions that take the place of referring expressions. Notice that they can all take the adjectival position:

(3a’) There are *so many/however many/that many* planets in the solar system.
(3b’) Jupiter has a mass of *so many/however many/that many* kilograms.

There are no rules taking us from (1a”–b”) and (3a’–b’) to sentences containing ‘The number twelve’ or ‘The number 1.8986×10^{27}’. This provides a good reason for thinking that ‘eight’ and ‘1.8986×10^{27}’ in (1a–b) are not behaving semantically as referring expressions, and, since ‘so many’ and ‘however many’ (and perhaps even ‘that many’) can plausibly be seen as standing in for determiners, this provides some reason for extending the GQT account to contexts such as (1a–b), and so analysing the numerals therein as determiners.

I have provided evidence in favour of a particular solution to the semantic dimension of FOP. The evidence at least suggests that the post-copular expressions in contexts such as (1a–b) are not referring expressions, and goes some way to justifying the claim that they occur as determiners. It appears that the Fregean analysis, put forward as a thesis about the semantics of natural language sentences, is mistaken. However, the syntactic dimension of FOP remains: why can non-referring expressions occur in singular-term position as well as their natural adjectival position? Before moving onto this question, I will now provide good reasons for thinking that the pre-copular expressions of sentences such as (1a–b) do not purport to refer to numbers, though they do purport to refer.

4.3 Number-of and Magnitude-of Expressions

On the Fregean analysis, number-of expressions and magnitude-of expressions stand for numbers. In this section, I examine evidence presented by Moltmann for the claim that number-of expressions stand instead for number tropes. Tropes are typically understood as particular instances of properties, including the beauty of a particular painting, or the fragility of a particular vase, and are the qualitative aspects of objects. Moltmann claims that number tropes are the quantitative aspects of pluralities.

Moltmann presents two kinds of evidence. The first is meant to show that number-of expressions are referring expressions that do not stand for numbers. To see that 'The number of planets in the solar system' is a singular term, consider:

(10a) The number of planets in the solar system is small.
(10b) The number of planets in the solar system is surprising.
(10c) The number of planets in the solar system is the same as the number of chickens.

We have little choice but to interpret 'The number of planets in the solar system' as standing for an object that each predicate is describing.

Moltmann argues that number-of expressions do not stand for numbers by demonstrating that constructions permissible with number-referring expressions are not permissible with number-of expressions, and vice versa (Moltmann 2013a, pp. 502–504). Consider:

(11a) * The number eight is small.
(11b) * The number eight is surprising.
(11c) * The number eight is the same as the number eight.
(11d) The number eight is the number eight.

Though (11c) may be permissible, it is less natural than (11d). (11a) and (11b), though grammatical, make little sense. Similarly, when we rewrite (11c) as a straightforward identity statement, we get:

(12a) * The number of planets is the number of chickens.

This sounds odd, if it makes sense. If the number of pigs and the number of chickens were numbers, we should expect such an identity statement to sound fine.

This leads Moltmann to conclude that number-of expressions do not stand for numbers. However, the evidence is not conclusive. The permissibility of (10a–c) shows that, in those contexts, 'The number of planets in the solar system' doesn't stand for a number: it is that there are eight planets that is surprising. It does not show that number-of expressions do not stand for numbers in other contexts. Suppose that John typically wants things that are large and non-shiny, but now he wants a diamond. Compare:

(13a) What John wants is unusual.
(13b) What John wants is small and shiny.

In (13a), it is that John's wants the diamond that is said to be unusual, the diamond itself need not be unusual for the sentence to be true. However, (13b) appears to describe the diamond. The same expression, 'What John wants', stands for a different object depending on the predicate used. Similarly, 'The number of planets in the solar system' does not stand for a number in (10a–c), but in certain mathematical contexts, it clearly does:

(14a) The number of planets in the solar system is even.
(14b) The number of planets in the solar system is odd.

Moltmann points out that certain mathematical predicates make unacceptable constructions:

(15a) * The number of planets in the solar system is rational.
(15b) * The number of planets in the solar system is real.

Moreover, she claims there is a unifying feature of mathematical contexts that are permissible: they can be verified or falsified by performing operations on collections of objects (Moltmann forthcoming). (14a) implies the planets can be divided into two groups of equal number; (14b) implies they can't. Both can be verified or falsified by arranging the objects and counting them. Moltmann claims this is explained by the fact that number-of expressions refer to aspects of collections.

However, there is an alternative explanation for the oddness of (15a–b). The adjectives 'real' and 'rational' are ambiguous between their mathematical and everyday meaning. The former is appropriate for describing numbers, while the latter is appropriate for describing worldly objects. If number-of expressions refer to different things depending on their adjoining predicate, then the strangeness of (15a–b) would be unsurprising: the referent of 'The number of planets in the solar system' depends on the interpretation of the predicate, and the interpretation of the predicate depends on the referent of the referring expression. To dispel the ambiguity, it must be made clear that the predicate is mathematical:

(16a) The number of planets in the solar system is a rational number.
(16b) The number of planets in the solar system is a real number.

This sounds fine because the predicate is unmistakably mathematical, forcing an interpretation of 'The number of planets in the solar system' as standing for a number. *Contra* Moltmann, it appears that number-of expressions do purport to stand for numbers in some mathematical contexts. Nevertheless, it is clear that in some non-mathematical contexts they stand for something else.

Moltmann's second kind of evidence is supposed to show that the referents of number-of expressions share properties with the relevant collections. Consider:

(17a) The number of women is unusual.
(17b) The women are unusual, in number.

The predicate 'is unusual' can be applied to both the number of women, and the collection of women, so long as the modifier 'in number' is added. Moltmann claims this shows that the collection of women and the number of women share the property of being unusual. She claims that, because ‘in number’ must be added to preserve the meaning of (17a), it is something to do with the quantitative aspect of the women that is unusual.

This evidence is dubious. It is true that the number of women bears a special relation to the collection of women that allows the move from (17a) to (17b), but this does not mean that these two entities share the property of being unusual. Compare:

(18a) The font of the book is large.
(18b) The book is large, in terms of its font.

The book and the font stand in a relation that permits the move from (18a) to (18b), but the book and the font do not share the property of being large. (18a) and (18b) can be true when the book is small. Similarly, it is the number of women that is unusual; the women need not be.

Suppose these two entities do share properties. It does not immediately follow that number tropes are the ideal candidates for the referents of number-of expressions. Moltmann argues that the kinds of properties shared by the referents of number-of expressions and collections can only be instantiated by concrete entities, and so concludes that the referents of these terms are number tropes, the concrete, quantitative aspects of collections (forthcoming).

The standard notion of concreteness is assumed: an entity is concrete iff it can act as an object of perception, a relatum of causal relations, and is spatio-temporally located. Moltmann points out that perceptual and causal predicates apply to the referents of number-of expressions:

(19a) John noticed the number of women.
(19b) The number of women caused Mary consternation.

This is poor evidence. Examples attributing causal and perceptual properties to things that cannot enter into causal or perceptual relations abound:

(20a) John began to see Mary’s point of view.
(20b) Fermat’s Last Theorem caused Mary frustration.

Objection: ‘see’ is figurative in (20a). But that is precisely the point. Perceptual predicates are often used in this way to express that something is understood, and there is no evidence suggesting that ‘noticed’ in (19a) is not used in this capacity. Similarly in (20b), though a theorem is not causal, we would say that Fermat’s Last Theorem causes frustration if attempts to prove it are frustrating. There is no evidence to rule out that ‘caused’ is used in this loose way in (19b).

Moltmann’s criterion for concreteness includes spatio-temporal location, and yet she admits that spatio-temporal predicates do not apply to the referents of number-of expressions (Moltmann 2013a, p. 505; Moltmann 2013b, pp. 56–57). Consider:

(21a) * The number of cats is in the bedroom.
(21b) * The cats are in the bedroom, in number.
(21c) * The cats are no longer, in number.
(21d) * The number of cats is no longer.

If number tropes were the concrete quantitative aspects of collections, they would go wherever and whenever the collections go. Even if the number of cats and the collection of cats share properties, there is no evidence that they share concrete properties. There is no evidence that number tropes are the referents of number-of expressions. What can we conclude about number-of expressions and magnitude-of expressions? We have seen that, in come subject-predicate constructions, they do not stand for numbers, while in others, they do. Recall:

(10b) The number of planets in the solar system is surprising.
(16b) The number of planets in the solar system is a real number.

In (10b) it is that there are eight planets that is surprising, while in (16b), it is a number that is said to be real. We have seen that similar changes of reference are exhibited by other terms. Recall:

(13a) What John wants is unusual.
(13b) What John wants is small and shiny.

In (13a) it is *that John wants a diamond* that is unusual, while in (13b), it is the diamond that is small and shiny. This is corroborated by the fact that the following intuitively express the same propositions as (13a) and (10b), respectively:

(22a) That John wants a diamond is unusual.
(22b) That there are eight planets in the solar system is surprising.

It is plausible that the referent of 'What John wants' and 'The number of planets in the solar system' in these contexts is whatever the referent of the corresponding that-clause is. That-clauses are typically thought to denote propositions, but the following suggests that in these contexts they refer to facts:

(23a) The fact that there are eight planets in the solar system is surprising.
(23b) The fact that John wants a diamond is unusual.
(23c) *The proposition that there are eight planets in the solar system is surprising.
(23d) *The proposition that John wants a diamond is unusual.

The evidence presented in this section suggests that, in applied contexts, number-of expressions stand for facts, and the same goes for magnitude-of expressions. However, in mathematical contexts, it appears they stand for numbers. The question now is whether or not sentences such as (1a–b) provide a mathematical or non-mathematical context. If the Fregean analysis is correct, then they provide a mathematical context, since they concern numbers. However, in the previous section, we saw that there is good reason for thinking that the post-copular expressions in (1a–b) are not referring expressions, so the Fregean analysis is mistaken. But that does not yet settle the issue at hand. I leave this issue aside for the time being and return to it in Sect. 4.5. For now, I turn to the evaluation of some recent syntactic analyses of sentences such as (1a–b) that promise to solve the syntactic dimension of FOP.

4.4 Specificational Sentences

Hofweber (2005, pp. 210–211) suggests that the unusual syntactic position of numerals in sentences such as (1a–b) is due to a widespread phenomenon called ‘focus’, whereby expressions are moved to unusual syntactic positions in order to emphasise certain aspects of the information expressed. Compare:

(24a) John swims quickly to shore.
(24b) The way John swims to shore is quickly.

Intuitively, (24a–b) have the same truth-conditions. However, (24a) presents the information neutrally, while (24b) emphasises the way in which John swims. To the question ‘How does John swim to shore?’, (24a–b) are both appropriate answers. To the question ‘Where does John swim to?’, (24a) is appropriate, while (24b) is not. This effect is focus.

Hofweber claims that sentences such as (1a–b) exhibit numerals in an unusual position to emphasise how many of the relevant objects there are. Indeed, to the question ‘What revolves around the sun?’ (1a) would be a strange answer, while ‘There are eight planets in the solar system’ seems perfectly appropriate. To ‘How many planets are in the solar system?’, both would be acceptable. Hofweber claims that, just like (24a–b), (1a) and (3a) have the same truth conditions.

However, this is not yet a solution to FOP. If (1a) and (3a) have the same truth-conditions, then ‘The number of planets in the solar system’ in (1a) cannot function as a referring expression. We saw in the previous section that it typically does. The unusual semantic behaviour of ‘The number of planets in the solar system’ in (1a) requires explanation on Hofweber’s account. He doesn’t offer one, so his solution is incomplete.

Felka (2014, pp. 263–265) points out that focus is characteristic of specificational sentences, so Hofweber’s examples should be taken as evidence that sentences such as (1a–b) are specificational sentences, in which case, the solution to the syntactic dimension of FOP requires an analysis of specificational sentences. Compare the following examples of predicational and specificational sentences (Mikkelsen 2011, p. 1806): Predicational:

(25a) The hat I bought for Harvey is big.
(25b) What I bought for Harvey is big.

Specificational:

(26a) The director of Anatomy of a Murder is Otto Preminger.
(26b) Who I met was Otto Preminger.

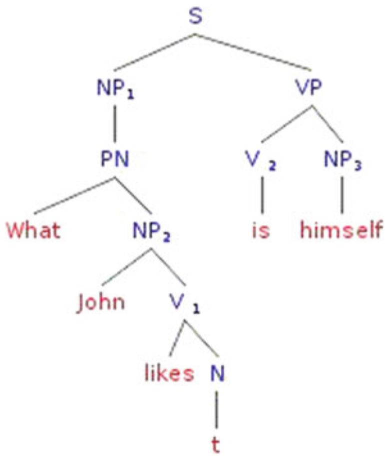
Examples of predication and specification share syntactic features. (25b) and (26b) begin with an interrogative pronoun as part of what is known as a ‘wh-clause’. Both are ‘pseudoclefts’. There are predicational and specificational pseudoclefts. (25a) and (26a) begin with a headed relative clause and are plain predication and plain specificational sentences, respectively. Linguists agree that the distinction between predication and specificational is the more semantically important one. It is therefore desirable to have a semantics of predication sentences that unifies all their syntactic forms, including plain and pseudocleft, and a semantics of specificational sentences that unifies all theirs (cf. Mikkelsen 2011, p. 1807; Felka 2014, p. 268). Specificational sentences are used to specify who or what something is: (26a) specifies who the director of a certain movie is and (26b) specifies who the speaker met. Specificational sentences also exhibit focus, giving emphasis to the post-copula expression. (1a–b) share these features so it is highly plausible that such sentences are specificational sentences. It is hard not to read (1a) as specifying how many planets there are, and (1b) as specifying how many kilograms the planet's mass is, and we have seen that these sentences exhibit focus. Some theorists (Moltmann forthcoming; Felka 2014; Schlenker 2003) have analysed specificational sentences as question-answer pairs: the pre-copula expression as an interrogative; the post-copula an elided answer clause:

(27a) What James likes is lying down.
(27b) [What James likes is?] is [James likes lying down].

The evidence in support of this analysis is that it solves a syntactic puzzle. Compare:

(28a) John likes himself.
(28b) What John likes is himself.

In both, 'himself' borrows its referent from 'John'. According to binding theory, this is only possible if the two expressions stand in a specific syntactic relationship: c-commanding. This is cashed out in terms of dominance, represented in syntax trees. A node *x* dominates a node *y* iff *x* is above *y* and one can trace a line from *x* to *y* while only moving downwards. An expression *a* c-commands for an expression *b* iff the first branching node dominating *a* also dominates *b*, and neither *a* nor *b* dominate each other. See the representation of (28b) below:



By the above structure, attributed to (28b) by binding theory, 'John' does not c-command 'himself'. Yet, according to all the usual tests, (28b) is a case of c-commanding. This is the puzzle of connectivity.

The question-answer analysis promises to solve this because it attributes a structure to specificational sentences that meets the criteria imposed by binding theory: 'himself' in (28b) is elliptical for the answer clause 'John likes himself', whose structure is identical to (28a). The pronoun therefore borrows from the hidden occurrence of 'John', which is in the right position to lend its reference.

There is more evidence in support of the question-answer analysis, the most impressive of which is the following. Compare:

(29a) What I did then was call the grocer.
(29b) What I did then was I called the grocer. (Ross 1972, (39a), (39b))

Here the ellipsis in the post-copula clause is optional. Sometimes, it is not, but Schlenker (2003, p. 14) proposes a plausible explanation for this. Consider:

(30c) Who I called was the grocer.
(30d) ? Who I called then was I called the grocer.

Schlenker suggests that the repetition of the verb makes ellipsis obligatory. In contrast, there is no danger of repetition of the verb in (29b). There are good reasons for adopting the question-answer analysis, but it is important to note that the most compelling evidence suggests only that the post-copular expression is an elided answer.

Moltmann (2013a, pp. 521–525) is a proponent of the question-answer analysis, and proposes to analyse arithmetical specificational sentences as follows:

(31a) The number of ship is seven.
(31b) [How many ships?] is [There are seven ships]. (forthcoming)

The unpacked form of 'seven' has the numeral occurring in its typical syntactic position. This fits with the semantic analysis of numerals presented in the previous section, and offers a deeper explanation of their unusual syntactic distribution: they occur in elided answer clauses in which they hold their standard position. However, Moltmann's analysis runs into a problem similar to that encountered by Hofweber. According to Moltmann, number-of expressions occur as referring expressions in most contexts, but disappear under analysis here. This phenomenon requires explanation. Moreover, no argument is offered for why the interrogative is 'How many ships?' rather than 'What is the number of ships?' The latter better resembles the original sentence, and I can think of no argument for preferring Moltmann's analysis.

It is also somewhat farfetched that apparent definite descriptions can be analysed as interrogatives at all: 'The number of ships' just doesn't seem like a question. Sensitive to this, Felka posits yet more ellipsis:

(32) [What the number of moons of Jupiter is] is [Jupiter has four moons]. (Felka 2014, 276)

Felka claims that there are other examples of embedded interrogatives where this ellipsis is optional:

(33) John revealed who the winner of the competition is.

This is an improvement on Moltmann's analysis, but it suffers from significant drawbacks. For one, it implies that definite descriptions and so-called interrogatives are semantically identical in certain contexts. Elsewhere, I present evidence to show that this implication is false, and that the syntactic distributions of these expressions differ in ways that the question-answer analysis cannot adequately explain (Knowles 2015, pp. 2766–2771). In this paper, however, I want to undermine the question-answer analysis on different grounds. In the following section, I present a semantics of interrogatives that suggests we should instead think of expressions like 'Who the winner of the competition is' as *answer clauses* rather than question clauses. Along with the evidence in favour of accepting that the post-copular expressions of specificational sentences are elided answers, this suggests an analysis of specificational sentences as answer-answer pairs, leading naturally to the semantics of specificational presented in Sect. 4.6.

4.5 The Semantics of Interrogatives

In their seminal paper, Groenendijk and Stokhof (1997) motivate a simple and explanatorily powerful semantics of interrogatives that emerges from three plausible principles:

(i) An answer to a question is a proposition.
(ii) The possible answers to a question are an exhaustive set of mutually exclusive possibilities.
(iii) To know the meaning of a question is to know what counts as an answer to it.

The first principle falls out of two plausible considerations: first, that the proper linguistic vehicle for an answer is a declarative sentence; second, that a question asks for a certain piece of information. The second principle puts a constraint on propositions that count as answers to a question: the truth of one entails that the rest are false, so they logically exclude one another—they are *exhaustive*. The set of answers to a question completely fills the logical space defined by the question. The picture that emerges is that the semantic value of an interrogative is a partition of logical space. The set of possible worlds at which the semantic presuppositions of the question are true is the logical space defined by the question and the cells of the partition are the subsets, each constituting a proposition that is an exhaustive possible answer to the question.

The final principle parallels one of the foundational principles of truth-conditional semantics of indicatives. Just as we assume that to know the meaning of an indicative is to know the conditions under which it is true, we assume that to know the meaning of an interrogative is to know the conditions under which it is answered. On this view, the intension of an interrogative is a partition, and its extension relative to a world is the proposition that is its unique and true answer at that world. Call this the 'partition analysis'.

As well as its intuitive appeal, the partition analysis has considerable explanatory power. It allows us to account for apparent meaning relations between interrogatives. For example, the interrogatives 'Who came to the party?' and 'Which women came to the party?' stand in a sort of entailment relation in that a complete answer to the former entails a complete answer the latter. On the partition analysis, this can be dealt with in familiar terms: question P? entails question Q? iff each cell of the partition denoted by P? is included in some cell of the partition denoted by Q?. The analysis also makes sense of cases in which an indirect interrogative is embedded in an intensional verb, such as 'wonder':

(34a) Cooper wondered who killed Laura Palmer.
(34b) * Cooper wondered that Bob killed Laura Palmer.

Because 'wonder' is an intensional verb, its semantic value takes the intension of the indirect interrogative as an argument. A partition is an unresolved entity, so to speak, and truth and falsity cannot be predicated of it, so it seems the appropriate object for the relation expressed by the verb. That 'wonder' selects for an object such as this can be seen from (34b): it cannot take a fact or proposition-denoting expression as argument because facts are 'resolved' and propositions bear truth values.

However, the partition analysis has trouble accounting for interrogatives embedded in non-intensional contexts. Consider the following invalid argument (some of the following is adapted from Ginzburg 1995):

(35a) Cooper discovered an interesting question.
(35b) The question he discovered was who killed Laura Palmer.
(35c) Therefore, Cooper discovered who killed Laura Palmer.

The partition-based explanation for the invalidity of this argument would presumably have to rest on the claim that 'Who killed Laura Palmer' in (35b) contributes its intension, a partition, while in (35c) it contributes its extension, namely whichever proposition constitutes its true and exhaustive answer at the relevant world. The problem arises when we test these claims directly. It turns out that indirect interrogatives cannot occupy positions that allow for proposition-denoting expressions:

(36a) Cooper believed an interesting proposition.
(36b) ? The proposition Cooper believed was who killed Laura Palmer.
(36c) ? Therefore, Cooper believed who killed Laura Palmer.

They do, however, occupy positions reserved for fact-denoting expressions:

(37a) Cooper discovered an interesting fact.
(37b) The fact Cooper discovered was who killed Laura Palmer.
(37c) Therefore, Cooper discovered who killed Laura Palmer.

Moreover, we have already seen in Sect. 4.3 that they serve as subject in subject-predicate constructions in which it is most plausible to assume that they denote a fact. For example:

(38) What John wants reflects badly on his character.

It is the fact that John wants ice cream that reflects badly on his character, not the proposition or the ice cream itself. We see the same patterns in nominalizations of indicative sentences, or that-clauses. Consider:

(39a) Cooper discovered an interesting fact.
(39b) The fact was that Bob killed Laura Palmer.
(39c) Therefore, Cooper discovered that Bob killed Laura Palmer.

Again, that-clauses can occur in subject-predicate constructions in which the most natural reading is that they stand for facts:

(40) That John likes ice cream reflects badly on his character.

Again, it is the fact that John likes ice cream that reflects badly on his character.

All of this suggests that we need an account of indirect interrogatives according to which, like that-clauses, they can refer to facts. Elsewhere, I have suggested one way of achieving this (Knowles 2015, pp. 2771–2773). If we understand facts as the parts of worlds that true propositions correspond to, then we can assign to each world the set of facts that form part of that world (the obtaining facts), and the set of facts that do not (the non-obtaining facts). Then we can assume that a sentence extensionally denotes the fact that makes it either true or false at that world, rather than its truth-value at that world, and treat indirect interrogatives as syntactically derived nominalizations of those sentences that refer to those facts. There is not space to present the proposal in detail here. Instead, I want to demonstrate that it yields an account of interrogatives that is both consistent with the motivations for principles (i)–(iii) above, and with the data concerning embedding in factive contexts.

In short, the view is as follows. Facts are assigned as the extension (and referent) of indirect interrogatives. From this, it follows that the intension of an indirect interrogative is a function from worlds to facts. Recall that the motivation for principle (i) was that questions are a request for information, and that the proper vehicle for an answer is an indicative sentence. Propositions are valuable for answering questions, but arguably not in and of themselves. Plausibly, they are well-suited to this purpose only insofar as they can tell us about what facts obtain in the actual world. Moreover, they are dispensable to this purpose. If I ask ‘Where are my keys?’, and then reach into my pocket and find them, I might say ‘That answers my question’. A more direct indication of the facts than via the communication of a true proposition can clearly provide answers to questions. It seems plausible, then, to take facts to be the primary answers to questions, and propositions merely as a convenient vehicle for representing them. As for the claim that sentences are the proper vehicles for answers, if facts are the extensions of sentences, then the most appropriate way of indicating the relevant fact will often be to utter the appropriate indicative. So, let us change (i)–(iii) accordingly:

(i) An answer to a question is a fact.
(ii) The possible answers to a question are an exhaustive set of mutually exclusive possibilities.

(iii) To know the meaning of a question is to know what counts as an answer to it.

This allows us to account for the occurrences of interrogatives embedded in factive verbs, such as ‘know’. On the present account, fact-denoting that-clauses and indirect interrogatives both denote facts. The validity of (37a–c) and (39a–c) is explained by the fact that these expressions share a denotation:

(37a) Cooper discovered an interesting fact.
(37b) The fact he discovered was who Killed Laura Palmer.
(37c) Therefore, Cooper discovered who killed Laura Palmer.

(39a) Cooper discovered an interesting fact.
(39b) The fact was that Bob killed Laura Palmer.
(39c) Therefore, Cooper discovered that Bob killed Laura Palmer.

When indirect interrogatives are embedded in intensional contexts such as ‘wonder’, it is their intension that is contributed. Recall that the object is required to be such that it is unresolved and cannot be true or false. A function from worlds to facts is unresolved in that it doesn’t settle the question of which fact actually obtains, and truth and falsity cannot be attributed to it, so the acceptability of (34a) is explained:

(34a) Cooper wondered who killed Laura Palmer.

Finally, propositions are not assigned as either the intensions or extensions of indirect interrogatives, so we expect that they would not be able to occupy positions reserved for proposition-denoting expressions. Recall:

(36a) Cooper believed an interesting proposition.
(36b) ? The proposition he believed was who killed Laura Palmer.
(36c) ? Cooper believed who killed Laura Palmer.

There is good reason to think that indirect indicatives refer to facts. There is one remaining issue that needs addressing, however. Any plausible analysis of interrogatives must explain the relationship between indirect interrogatives, such as ‘Where my keys are’, and direct interrogatives, such as ‘Where are my keys?’ (cf. Karttunen 1977, p. 3–4). On the partition analysis, partitions are assigned as the semantic values of both direct and indirect interrogatives, and I want to maintain this assignment with respect to the former. That a question divides logical space into propositions that represent the ways the world would have to be for the question to be answered is intuitive. In light of this, however, I must relinquish the assumption that expressions such as ‘Where my keys are’ are interrogatives at all. Instead, I propose we understand them as answer clauses. By an ‘answer clause’, I mean an expression that refers to the fact that answers the corresponding question. There are two good independent reasons for adopting this view. First, indirect interrogatives *sound* like answers. One can imagine a particularly facetious person providing (41b) as an answer to (41a):

(41a) What does John want?
(41b) What John wants.

It is not a useful way of communicating the answer, but it does point to the right fact. The second reason is that the account provides an elegant explanation of the relationship between indirect and direct interrogatives while still accounting for their difference in structure. Indirect interrogatives denote the facts that answer the questions posed by direct interrogatives; they differ in structure because they belong to different semantic types. I am now in a position to present my semantic account of specificational sentences.

4.6 Solution and Conclusions

My account of indirect interrogatives as answer clauses that refer to facts leads naturally to a plausible analysis of specificational sentences as pairs of answer clauses. We have seen that there is good reason to think that the pre-copular expressions of specificational sentences refer to facts. If we add to that the claim that the elided indicative answer clause in post-copular position also refers to the fact that makes the proposition it expresses true, then we get an analysis according to which specificational sentences identify facts. Call this the 'fact analysis' (FA).

I can now state the truth-conditions of a specificational sentence. (42b) indicates the syntactic analysis, while (42c) specifies the truth-conditions:

(42a) The one who knocks is Walter White.
(42b) [The one who knocks] is [Walter White knocks]
(42c) 'The one who knocks is Walter White' is true iff: the fact that uniquely and exhaustively answers the question 'Who is the one who knocks?' is identical with the fact that Walter White knocks.

Perhaps more needs to be said about what I mean by facts. I do not have the space or the inclination to provide a metaphysical account of facts here. In this context, they are to be understood primarily as the referents of fact-referring expressions. I suggested in the previous section that we think of them as the parts of worlds in that true propositions correspond to. So, a proposition true at a world corresponds to a fact that is part of that world, and a proposition that is false at that world corresponds to a fact that is not part of that world, but part of some other world. If (42a) is true, the fact that Walter White knocks is the part of the world in virtue of which the proposition that Walter White knocks is true. Moreover, if (42a) is true, the fact that Walter White knocks is also the part of the world that uniquely and exhaustively settles the question 'Who is the one who knocks?'. This happens just in case Walter White is the only one who knocks. Intuitively, then, FA gets the truth-conditions for (42a) right.

Doesn't the answer-answer analysis run counter to the evidence presented in favour of the question-answer analysis presented in Sect. 4.4? No. The evidence suggests only that the post-copular clause is an elided answer, and FA is consistent with this. Schlenker (2003) does provide evidence that he takes to suggest that the pre-copular expressions of specificational sentences are interrogatives, but it all presupposes that expressions like ‘Where my keys are’ are interrogative clauses when they occur in other, non-specificational contexts. In the previous section, I provided good reasons for thinking that such expressions are fact-referring answer clauses across the board. For this reason, I do not find Schlenker’s evidence compelling.

Though they are different, FA is similar to the analysis of specificational sentences defended by Schlenker (2003), who also takes them to be identity statements. He is a proponent of the question-answer analysis and the partition view of interrogatives. He claims that specificational sentences are identity statements that equate the proposition that, according to him, amounts to the exhaustive unique true answer to the relevant question. We have seen the problems that the partition view faces. In particular, it had problems explaining the occurrence of indirect interrogatives in factive contexts, and their failure to occur in positions reserved for proposition-denoting expressions. These problems extend to cases involving specificational sentences. On the one hand, Schlenker has difficulty explaining the validity of arguments such as the following:

(43a) Mary discovered an interesting fact, namely what John had for pudding last night.
(43b) What John had for pudding last night is ice cream.
(43c) Therefore, Mary discovered that John had ice cream for pudding last night.

The validity of this argument seems to require that facts are referred to throughout. On Schlenker’s account, however, it is difficult to makes sense of (42a), and more generally to make sense of the validity of the above argument. On the other hand, Schlenker’s account also has problems accounting for the infelicity of the following:

(44a) *Mary believed an interesting proposition, namely what John had for pudding last night.
(44b) What John had for pudding last night is ice cream.
(44c) Mary believed that John had ice cream for pudding last night.

On Schlenker’s view, the above should be acceptable, but it isn’t. This is to be expected on FA, since propositions are not assigned to either the pre- or post-copular expressions of specificational sentences (see Knowles 2015, pp. 2772–2773 for more reasons to prefer FA to Schlenker’s view).

I will now state the truth-conditions of (1a–b) and discuss what we can conclude from them with regards the arguments mentioned in the introduction.

(1a’’’) ‘The number of planets in the solar system is eight.’ is true iff: the fact that uniquely and exhaustively answers the question ‘How many planets are there in the solar system?’/‘What is the number of planets in the solar system?’ is identical to the fact that there are eight planets in the solar system.
(1b’’’) ‘The mass of Jupiter in kilograms is 1.8986×10^{27}.’ is true iff: the fact that uniquely and exhaustively answers the question ‘What is the mass of Jupiter in kilograms?’ is identical to the fact that the mass of Jupiter in kilograms is 1.8986×10^{27}.

The first thing to note here is that the truth-conditions of (1a–b) do not mention mathematical objects. They mention certain kinds of facts. In the case of sentences such as (1a), what we might call ‘cardinality facts’; in cases such as (1b), what we might call ‘magnitude facts’. So, the Fregean analysis does not provide an adequate account of the semantics of these sentences, and the truth-conditions of sentences such as (1a–b) do not appear to involve mathematical objects. On the face of it, then, proponents of easy arguments and indispensability arguments cannot appeal to applied arithmetical language.

Is there a way for the realist to salvage these arguments? There are two routes to take. The first is to focus on different kinds of mathematical sentences. This will be less successful in the case of easy arguments. Easy arguments are effective because they start from a highly intuitive premise, that sentences such as (1a–b) are true. If the realist were to instead appeal to pure mathematical language in support of an easy argument, I suspect that intuitions would not be as strongly in their favour. For the proponent of the indispensability argument, however, appealing to the applications of pure mathematical language in science is more promising. For instance, pure arithmetical language is intuitively about numbers, and enjoys many applications in science. However, if this paper has shown anything, it is that semantic assumptions such as this should not be taken lightly. A successful indispensability argument would need, inter alia, to provide evidence in favour of its semantic assumptions.

The other route is instead to maintain on metaphysical grounds that the truth of applied arithmetical sentences, such as (1a–b), does require mathematical objects to exist. Granted, the truth-conditions of (1a–b) only mention cardinality facts and magnitude facts; but they do not say anything about the nature of these facts. The realist could argue that, in order for such facts to obtain, there must be mathematical objects. The view that magnitude facts consist in physical objects bearing relations to numbers has a name, ‘Heavy Duty Platonism’, and might naturally be thought to go hand in hand with the corresponding view about cardinality facts. Heavy Duty Platonism is not a popular view, but, recently, I (Knowles forthcoming) have shown that all the arguments against it presented and alluded to in the literature fail. Again, adding a defence of Heavy Duty Platonism to an easy argument will render it no longer worthy of its name, but defending Heavy Duty Platonism may be a promising means for the realist to defend an indispensability argument that appeals to applied arithmetical language.

References

Barwise, F., & Cooper, R. (1981). Generalised quantifiers and natural language. *Linguistics and Philosophy*, *4*, 159–219.

Dolby, D. (2009). The reference principle: A defence. *Analysis*, *69*, 286–296.

Felka, K. (2014). Number words and reference to numbers. *Philosophical Studies*, *168*, 261–282.

Field, H. (1980). *Science without numbers: A defence of nominalism*. New Jersey: Princeton University Press.

Field, H. (1989). *Realism, mathematics and modality*. Oxford: Blackwell.

Frege, G. (1884). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl. Breslau*: W. Koebner. English edition: Frege, G. 1953. *The foundations of arithmetic: A logico-mathematical enquiry into the concept of number* (J. Austin, Trans). Oxford: Blackwell.
Ginzburg, J. (1995). Resolving questions II. *Linguistics and Philosophy*, *18*, 567–609.
Groenendijk, J., & Stokhof, M. (1997). Questions. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 1055–1124). Cambridge: MIT Press.
Hofweber, T. (2005). Number determiners, numbers, and arithmetic. *Philosophical Review*, *114*, 179–225.
Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*, 3–44.
Keenen, L., & Westerståhl, D. (1997). Generalized quantifiers in linguistics and logic. In J. Benthem & A. Meulen (Eds.), *Handbook of logic and language*. Cambridge: MIT Press.
Knowles, R. (2015). What 'the number of planets is eight' means. *Philosophical Studies*, *172*, 2757–2775.
Knowles, R. (Forthcoming). Heavy duty Platonism. *Erkenntnis*.
Künne, W. (2006). Properties in abundance. In P. F. Strawson (Ed.), *Universals, concepts, and qualities: New essays on the meaning of predicates and abstract entities* (pp. 249–300) Aldershot: Ashgate.
Mikkelsen, L. (2011). Copular clauses. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 2, pp. 1805–1829). Berlin: Mouton de Gruyter.
Moltmann, F. (Forthcoming). The number of planets, a number-referring term? In P. Ebert & M. Rossberg (Eds.), *Abstractionism*. Oxford: Oxford University Press.
Moltmann, F. (2013a). Reference to numbers in natural language. *Philosophical Studies*, *162*, 499–536.
Moltmann, F. (2013b). *Abstract objects and the semantics of natural language*. Oxford: Oxford University Press.
Montague, R. (1974). The proper treatment of quantification in ordinary english. In R. Thomason (Ed.), *Formal philosophy*. New Haven: Yale University Press.
Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, *44*, 12–36.
Oliver, A. (2005). The reference principle. *Analysis*, *65*, 177–187.
Prior, A. N. (1971). *Objects of thought*. Oxford: Clarendon Press.
Quine, W. V. (1948). On what there is. *The Review of Metaphysics*, *2*, 21–38.
Rosefeldt, T. (2008). 'That'-clauses and non-nominal quantification. *Philosophical Studies*, *137*, 301–333.
Ross, J. (1972). Act. In D. Davidson & G. Harman (Eds.), *Semantics of natural language*. Dordrecht: D. Reidel and Company.
Schlenker, P. (2003). Clausal equations: A note on the connectivity problem. *Natural Language and Linguistic Theory*, *21*, 157–214.
Sellars, W. (1960). Grammar and existence: A preface to ontology. *Mind*, *69*, 499–533.

Chapter 5
The Modal Status of Contextually A Priori Arithmetical Truths

Markus Pantsar

Abstract In Pantsar (2014), an outline for an empirically feasible epistemological theory of arithmetic is presented. According to that theory, arithmetical knowledge is based on biological primitives but in the resulting empirical context develops an essentially a priori character. Such contextual a priori theory of arithmetical knowledge can explain two of the three characteristics that are usually associated with mathematical knowledge: that it appears to be a priori and objective. In this paper it is argued that it can also explain the third one: why arithmetical knowledge appears to be necessary. A Kripkean analysis of necessity is used as an example to show that a proper analysis of the relevant possible worlds can explain arithmetical necessity in a sufficiently strong form.

Keywords Epistemology of arithmetic · Modality of arithmetical truths · Contextual a priori

5.1 Introduction

At first sight, arithmetical truths appear to have two characteristics that an epistemological theory should be able to explain: they seem to be both *objective* and *necessary*. Whether or not this is the case under deeper philosophical analysis, it can hardly be denied that the impression is strong. Although there are obviously exceptions, generally mathematicians, philosophers and laymen all find it hard to believe that, say, "$2 + 1 = 3$" is merely some contract we have agreed upon. They may disagree on what numbers are, or what it means for a truth to be objective, but in some relevant sense they nevertheless regard $2 + 1 = 3$ as an objective truth about numbers. But if it indeed is an objective truth, it would also seem to be a necessary truth. If someone is not ready to accept that "$2 + 1 = 3$" is merely a convention, she is not likely to accept that it could have been the case that $2 + 1 = 3$ was *not* true.

M. Pantsar (✉)
University of Helsinki, Helsinki, Finland
e-mail: markus.pantsar@gmail.com

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_5

In addition to those two impressions, it is common to have the impression that arithmetical knowledge is a priori.[1] It could be said that these three characteristics form the traditional (Kant 1781) understanding of the nature of arithmetical knowledge.[2] While this understanding is widely accepted, it has also been challenged from many directions. Conventionalists like Wittgenstein (1956), for example, deny the objectivity of arithmetical truths, while empiricists like Kitcher (1983) reject the a priori nature.

However, there is something potentially troubling in such challenges if they simply reject the impressions that we should be able to *explain*. The fictionalist Field (1980), for example, argues that the strong impression we have of the truth of certain mathematical theories is mistaken:

> What makes the mathematical theories we accept better than these alternatives to them is not that they are true [...] but rather that they are more useful. [...] Thus mathematics is in a sense empirical but only in the rather Pickwickian sense that it is an empirical question as to which mathematical theory is useful. (Field (1980), p. 15. Italics in the original)

But surely we cannot be satisfied to simply leave the matter there. In an account like Field's, the usefulness of some mathematical theories over their alternatives would seem to be one of the key philosophical questions. And just like we must ask why certain theories are useful while others are not, we must ask why the traditional image of arithmetic has been so appealing to mathematicians and philosophers—as well as other scientists and laymen. What is it about arithmetic that makes so many see it as an a priori pursuit of objective and necessary truths? Certainly the reason does not seem to be lack of knowledge about arithmetic, since so many experts subscribe to that view. In philosophy we may argue that arithmetic is neither a priori, objective nor necessary, but even in rejecting all those characteristics we cannot escape the question why it intuitively seems to us to have these characteristics. Ultimately, any epistemological theory of arithmetic should be able to deal with this problem.

In this paper I will study these questions with regard to the *contextual* a priori theory of epistemology arithmetic proposed in Pantsar (2014). In that account, arithmetical knowledge is thought to be empirically constrained by biological primitives that we have as infants and share with many animals. However, in that empirical context arithmetic is thought to be essentially a priori. I will first show that the appearances of a priori and objectiveness follow directly from that theory. The situation is more difficult, however, when it comes to necessity. Since sufficiently developed biological organisms do not develop in all possible worlds, the conception of necessary truths as being true in all possible worlds seems like a bad fit with the contextual a priori account. I will argue, however, that by limiting ourselves to the relevant set of possible worlds—i.e., ones with sufficiently developed organisms—we get a strong

[1] It is not possible here to enter the discussion of what—if anything—constitutes a priori knowledge. In this paper, I will accept the possibility of a priori knowledge and follow the classical understanding that it is knowledge that can be gained in an essentially non-empirical manner.

[2] For Kant, there was of course also the important distinction between synthetic and analytic knowledge. In this paper that distinction, namely the possible differences between arithmetical knowledge being *analytic* a priori or *synthetic* a priori, will not be considered.

enough conception of necessity. I will argue that this is consistent with a Kripke-esque theory of designation, when qualified for the relevant context of arithmetic. However, the approach is not limited to any particular semantic theory, as it only involves considering necessity in the proper context of arithmetic understood as an ability arising from our cognitive architecture.

5.2 Contextual A Priori

In Pantsar (2014), I have proposed an empirically feasible approach to the epistemology of arithmetic. The purpose of the paper is to study what kind of epistemological theory is required if the recent results in neurobiology, cognitive science and psychology are accepted as giving insight to the origins and nature of arithmetical knowledge. Those results suggest that arithmetical knowledge is built on biological primitives that we already possess as infants and share with many nonhuman animals. In short, the brain is structured so that there is a natural tendency to process observations in terms of quantities. This can be seen in subitizing, i.e. the ability to determine the quantity of objects in one's field of vision without counting, as well as in the capacity to estimate quantities and keep them in the working memory.[3]

This primitive ability, what I call *proto-arithmetic*, should not be confused with developed arithmetical thinking. The proto-arithmetical abilities quickly lose their accuracy as the numerosities become larger, and the quantities can be properly described as discrete only for the numerosities from one to four or five.[4] However, it is argued that this primitive ability to deal with numerosities works as the basis on which we develop our arithmetic, in a process in which the development of language plays a key role. The central idea is that the proto-arithmetical treatment of numerosities develops into actual arithmetic once explicit rules and number words or symbols are introduced. In developing arithmetical ability, one key stage is understanding some form of successor operation. The proto-arithmetical cognitive

[3]The empirical details and references can be found in Pantsar (2014). For an introduction to the empirical studies, see Dehaene (2011) and Dehaene and Brannon (eds) (2011). In the latter, the paper by Nieder (2011) is particularly recommended as a good window to the state of the art in the empirical research of proto-arithmetical cognition.

[4]Hence at the level of proto-arithmetic, it is better to talk of *numerosities* rather than *numbers*. This distinction is too rarely made in the empirical literature. Unfortunately, it is also common in the empirical literature to postulate needlessly strong abilities to the test subjects. For example, in one of the most famous papers, Wynn (1992), it is shown that infants react to unnatural numerosities in experimental settings. When they see one and one dolls put behind a screen but only one doll when the screen is removed, they are surprised. This was an important result and further experiments have shown the proto-arithmetical ability involved in the process to be a particularly important one. The unnatural numerosity actually surprises infants more than changes in the position, sizes, or even the character of the doll. Nevertheless, when Wynn called her paper "Addition and subtraction by human infants", she seems to postulate needlessly strong ability to her test subject. The infants could simply have the ability to hold one numerosity in their working memory and be surprised when the observations don't match that.

system enables understanding that numerals "one", "two", "three", and "four" refer to different numerosities that form a succession. In learning to count, we generalize on this process, that is, we understand that the step from "three" to "four" is no different from the step from "four" to "five", etc. (Carey and Sarnecka 2006; Spelke 2011; Pantsar 2014).

The final step of this development is formal arithmetic, which has traditionally been the subject of philosophy of arithmetic. But to understand arithmetic only as dealing with formal systems seems needlessly limiting, especially epistemologically. A key feature of my account is to see arithmetical knowledge as a subject involving different levels of ability with numerosities. Only then, I claim, can we fully understand the nature of arithmetical knowledge.

Thus characterized, arithmetical knowledge has minimal ontological and epistemological requirements. No mind-independent existence of arithmetical objects is assumed and no special epistemic faculty for arithmetical knowledge is needed—but neither is either of those denied. My main claim is that such a development from the proto-arithmetical ability based on biological primitives to the language-based arithmetical thinking is *enough* to explain the nature of arithmetical knowledge—including the impression we have of it as being a priori, objective and necessary.

In my outline for an epistemological theory for arithmetic, I characterize arithmetical knowledge thus understood as *contextually* a priori. The main idea behind that characterization is that arithmetical knowledge is indeed essentially a priori in character, but only in a context set by empirical facts. We do not accept the statement "$2 + 1 = 3$" because it can be derived from a set of arithmetical axioms which are knowable a priori, or because we have a special non-sensory arithmetical intuition. We accept it initially because by the time we have learned the necessary number words or symbols, we already have a vast experience of processing observations in terms of quantities. Our brain structure forces upon us the notion that one object added to two objects make three objects. From the point of view of a child learning about addition, she initially accepts "$2 + 1 = 3$" because it counts as an empirical *fact*—it corresponds to how she has experienced the world all her life.

However, after this empirical context is set, arithmetic does not work like empirical sciences. It has its own methodology of proof procedures (both formal and informal) that does not rely on empirical corroboration. While the roots of arithmetic are in empirical aspects, in that empirical context arithmetical knowledge is essentially a priori.[5]

[5] My epistemological theory is related to, but different from, other accounts emphasizing the empirical foundations of arithmetic, e.g. Kitcher (1983), Lakoff and Núñez (2000) and Jenkins (2008). In all the accounts perhaps the key question is how we can combine the position that arithmetical concepts are empirically grounded with the image of arithmetic as an a priori pursuit. The solutions range from rejecting the apriority (Kitcher) to accepting that arithmetical knowledge is essentially both empirical and a priori (Jenkins). One key feature of my account is to soften the notion of a priori applicable to arithmetic. However, it should be noted that the account of contextual a priori is fundamentally different from the ones by Kuhn (1993) and Putnam (1976). For me, arithmetic is not simply a paradigm that is a priori in one context but could be overthrown when

It is important to note that in the contextual a priori account, the empirical aspects do not play a merely *enabling* role for arithmetical knowledge. It is universally accepted that empirical methods, like manipulating collections of pebbles, help us learn arithmetic. But Frege (1884) famously distinguished those psychological processes (the context of discovery) from what makes arithmetical propositions *true* (the context of justification). In the beginning of *Begriffsschrift* (1879) he presents this distinction:

> we can inquire, on the one hand, how we have gradually arrived at a given proposition and, on the other, how it is finally to be most securely grounded. The first question may have to be answered differently for different persons; the second is more definite, and the answer to it is connected with the inner nature of the proposition considered.

When arguing that arithmetical knowledge is based on empirical aspects, am I mixing up the two contexts? Clearly this is not the case. While I take empirical aspects to be at the basis of arithmetical knowledge, there is an obvious difference in how an individual arrives at a given proposition (the discovery) and what the content of that proposition is. In studying the empirical roots or arithmetic, I am concerned with the inner nature of the proposition—I only claim that this inner nature cannot be fully explained in philosophy without tracking it back to its cognitive roots.

In the contextual a priori account, when we establish arithmetical truths, we do not engage in an empirical pursuit. The justification comes, just as for Frege, from an essentially a priori process. The contextual a priori account does not make arithmetic an empirical science that allows psychological processes to take over logical analysis as the justification of arithmetical statements. For all *practical* intents and purposes, arithmetic in the contextual a priori account is the same a priori discipline as it was for Frege.

The difference comes when we consider *why* arithmetical truths are what they are. This has traditionally been perhaps the most important question in the philosophy of mathematics and the answers have ranged from a platonic world of mathematical objects to mathematics being an empirical science. The contextual a priori account is one answer to that question. It states that arithmetical truths are ultimately constrained by the proto-arithmetical structure of our observations, determined by our cognitive architecture. It is that context that gives us the conception that there are such things as discrete quantities and they form a succession, and it is in that context that we justify arithmetical truths with logical methods.

Thus characterized, the contextual a priori account gives a direct explanation for one of the three impressions we have about arithmetical knowledge: that it is a priori in character. Once the empirical context is set—which happens early on in our lives—arithmetic *is* essentially a priori.

In addition to apriority, the account also gives a satisfactory explanation for the apparent objectivity of arithmetical truths. If arithmetic is based in such a way on

(Footnote 5 continued)
that context changes. Rather, arithmetical knowledge is based on biological primitives and as such it is based on an effectively inevitable way of experiencing the world.

biological primitives, arithmetical knowledge is based on an effectively inevitable way of experiencing the world. I argue that this gives arithmetical knowledge a sufficiently strong form of objectivity. We cannot help experiencing the world in terms of quantities, and this tendency is the basis of arithmetic. Perhaps we would like to call this “maximal intersubjectivity” rather than objectivity, but the important point is that arithmetical knowledge does not succumb to the conventionalist threat. Arithmetical truths—unlike, say, the rules of chess—are not only a matter of commonly agreed contract, and as such, we would expect them to appear objective to us.

In this way, the account for arithmetical knowledge in Pantsar (2014) seems to be able to explain two of the three main characteristics of the traditional image of arithmetical knowledge. Although arithmetical knowledge is not strictly a priori, it is contextually so. Moreover, this context is set by biological primitives, which explains why arithmetical truths appear to be objective to us. For someone dealing with arithmetic—from simple calculations to proving theorems—it is quite understandable to get the image she is engaging in a priori pursuit of objective truths. For all practical purposes, she is doing just that—in a context that she is thoroughly familiar with from years of experience in processing observations in terms of numerosities.

5.3 Necessity

While objectivity and the a priori appearance of arithmetic seem be to explained in sufficiently strong forms in the contextual a priori account, it is not clear that the impression that arithmetical knowledge is necessary can be explained as easily. In the tradition of analytic philosophy, a priori and necessary truths have often been thought to amount to the same thing, that is, they are grounded only in the *meaning* of the terms involved.[6] Clearly such a connection picks out an important class of sentences, those of the type “all bachelors are unmarried”.[7] But I have argued that arithmetical statements do *not* belong to that class, that is, they are not *strictly* a priori. In the traditional account of conflating a priori and necessary knowledge, it would seem to follow that arithmetical truths cannot be necessary, either.

When discussing modal concepts, arithmetical statements like “$2 + 1 = 3$” are often given as standard examples of necessary truths–that is how obvious the necessity of arithmetical truths is generally taken to be in philosophy. But if we understand necessary truths—as is usually done since Kripke (1963) and Lewis (1970), among others—as true in all possible worlds, arithmetical truths according to the contextual a priori account would *not* be necessary. There are possible worlds in which human

[6]Kripke (1980) p. 35 states that in contemporary (meaning around the year 1970) discussion very few philosophers even made the Kantian distinction between a priori and necessary.

[7]Here I am ignoring what I see as trivial empirical elements, like the fact that we need to see or hear (or feel) an explanation of the words involved in order to see the truth of sentences like “all bachelors are unmarried”.

beings and other animals with proto-arithmetical ability did not develop. If arithmetical truths are based on the proto-arithmetical ability, it would appear to follow that they are not necessarily true in such possible worlds. That is a quite troubling prospect. Most of us would not be ready to accept that "$2 + 1 = 3$" could be false, yet that would seem to be a possibility if we accept the contextual a priori account.

Are there ways out of this problem? One potential solution could be to think of arithmetical knowledge as arising from the proto-arithmetical ability, but arithmetical *statements* being true regardless of there being agents with knowledge of them. This way we could avoid, for example, the uncomfortable position that "$2 + 1 = 3$" could cease to be an arithmetical truth if there no longer were sufficiently developed biological organisms. But what would be the ontological status of arithmetical truths in such a scenario? Any answer would seem to come worryingly close to platonism or some other position that takes arithmetical truths to be totally mind-independent, which puts the whole point of the contextual a priori theory in question. One of the main strengths of the theory is its lack of ontologically problematic assumptions, in particular that of a mind-independent reality of mathematical objects. While certainly not contradictory with the contextual a priori, assuming a mind-independent existence of arithmetical truths at the very least goes against the general spirit of the theory.

Another possibility could be to accept that "$2 + 1 = 3$" may not be a truth in all possible worlds, but insist that this does not imply that it could be *false*. It could be the case that even primitive forms of arithmetic never emerged—perhaps because no sufficiently developed biological organisms developed—but if it did, it would have been the case that "$2 + 1 = 3$" is true. In such a scenario, either "$2 + 1 = 3$" is true, or there are no arithmetical truths. But also this solution seems to make an ontological claim that is too strong for the general spirit of the contextual a priori theory. According to the theory, arithmetic is based on our cognitive architecture. But it is possible that the brain structure could evolve in a sufficiently different manner to give birth to a radically different kind of arithmetic. To think that arithmetical knowledge must develop among the lines it has in our actual world suggests a stronger ontological status for arithmetic—that arithmetical truths somehow determine the way biological organisms must develop.

5.4 What Are Arithmetical Truths?

Based on the considerations above, explaining the apparent necessity of arithmetical truths appears to be a difficult task for the contextual a priori theory. The main obstacle seems to be combining an account that takes arithmetical knowledge to be based on biological primitives with the position that arithmetical truths must be true also in possible worlds where no such biological organisms exist. The big question is, however, how much weight we should give to such possible worlds. What if we instead limited our focus only to possible worlds with sufficiently developed biological organisms?[8]

Such a move may sound unwarranted, but I claim that it is because in philosophy we are still too committed to the traditional conception of arithmetic as a priori necessary truths. We use "$2 + 1 = 3$" as an example of a necessary truth because it is thought to give a prime example of a statement that is true regardless of the characteristics that possible worlds may have. But inherent in that is a strong form of objectivism about arithmetical truths. It may be easy to accept, for example, the position that arithmetical truths are eternal. However, it is not as easy to give that position a satisfactory explanation without evoking strong ontological commitments.

Perhaps we need to have a platonist conception that arithmetical truths concern mind-independent mathematical objects, which is hardly a satisfactory explanation for the contextual a priori theory. Alternatively, we may believe that arithmetical truths are actually *logical* truths. Yablo (2002) has argued for the position that arithmetic arises from our cognitive mechanisms, but that arithmetical facts are actually facts of first-order logic. Since logical truths are commonly accepted as really being necessary, this explains why arithmetical truths appear to be necessary.

However, it is not obvious how moving the focus to logic adds explanatory value. First of all, while we certainly can present finite quantifier-free arithmetical truths like "$2 + 1 = 3$" as truths of first-order logic, there is no generally accepted way of deriving full arithmetic purely from logic.[9] But even if we could derive full arithmetic from logic, would it explain anything further? Yablo argues that logical truths are indeed necessary since they are tautologies. It is not clear, however, why logical truths should be somehow easier to accept as necessary than arithmetical truths. Indeed, the necessity of "$1 + 1 = 2$" seems to be just as obvious as that of logical tautologies. Yablo explains this by claiming that arithmetical truths *are* in fact logical truths, and thus equally necessary. An alternative explanation would be that both simple arithmetical and simple logical truths are enforced upon us by our cognitive architecture, thus making both contextually a priori in character. Without further arguments, it is hard to see how we can distinguish between the apparently different natures of logical and arithmetical truths.

One such further argument has come from Maddy (2014), who also holds that simple arithmetical statements are shorthand for logical truths. Her account differs from Yablo's in that she allows that we need something more for the "..." part in the progression of natural numbers, i.e. when moving from finite numbers to the infinity of them. For this, she argues—following Bloom (2000)—that:

[8]By "sufficiently developed", I mean sufficiently developed along the lines that evolution took in our actual world. I do not wish to engage here in speculation about highly developed biological organisms with completely different cognitive systems.

[9]The neo-Fregeanism of Wright (1983) comes perhaps closest to fulfilling the logicist idea but in addition to using second-order logic, which is sometimes seen as going against the logicist ideal, he uses a non-logical axiom—the so-called *Hume's principle*—in deriving the axioms of arithmetic. Yablo's own solution is to interpret universal quantifiers in arithmetic as infinite chains of conjunctions.

> Much as our primitive cognitive architecture, designed to detect [the logical structure of the world], produces our firm conviction in simple cases of rudimentary logic, our human language-learning device produces a comparably unwavering confidence in this potentially infinite pattern. (Maddy 2014, p. 234)

Thus Maddy's theory is similar to the contextually a priori account when it comes the move beyond the primitive origins. However, when it comes to those primitive origins, her account has two differences. First of all, she holds that the cognitive architecture is about logical structure, i.e. it is proto-logical rather than proto-arithmetical. At this point, it is impossible to commit strongly to either position based on the empirical data. However, for the present matter at hand, it seems irrelevant whether our primitive abilities are primarily logical or arithmetical—or perhaps equally both. If they shape our experiences and thus constrain the content of arithmetic, the contextual a priori is equally applicable to both explanations.

The second difference is more relevant. In the above quotation, Maddy writes that our primitive cognitive architecture *detects* the logical structure of the world. In the contextual a priori account, the cognitive abilities are only thought to categorize our observations in terms of numerosities. This is clearly an important difference, since for Maddy rudimentary arithmetic (arithmetic without the "..." part) comes from the structure of the world, thus proposing stronger objectivism than the contextual a priori account.

This comes down to a version of the old Kantian problem of distinguishing between objective features of the world and those imposed by our cognitive architecture. Are we designed to *detect* the logical structure of the world or does our cognitive architecture *impose* that structure on our observations? Unfortunately, here I cannot go deeper into such considerations. But ontological carefulness being at the heart of the contextual a priori project, I am not prepared to accept Maddy's strong objectivist claim. In any case, her position is in no way incompatible with the contextual a priori account. Perhaps arithmetic *is* ultimately about objective features of the world and as such arithmetical truths are mind-independent and necessary in a stronger sense. What I am arguing for here is that even without making that assumption—for which I do not see enough grounds—we can have a perfectly satisfactory account of the apparent necessity of arithmetical truths.

This is an important point, because it seems that augmenting the contextual a priori theory of arithmetical knowledge with a strong form of objectivism would take away power from my argumentation. Even if the theory were essentially correct, we would now need to ask how the biological primitives are connected to either the platonic world of mathematical objects or the logical structure that all possible worlds share—which are just the kind of problems that the contextual a priori account was supposed to avoid.

5.5 Necessary Contextual A Priori Knowledge

It seems to be the case, however, that the reason arithmetical truths are such popular examples of necessary truths is not simply because they appear to be true in all possible worlds. Instead, I claim that impression is based on another impression: that arithmetical truths are a priori. If we can find out a truth simply by reflection, it is easy to accept that the truth in question holds in all possible worlds. But of course I have argued that we do *not* find out arithmetical truths purely a priori. Consequently, we should also be prepared to reconsider our understanding of arithmetical necessity.

In modern analytic philosophy it is widely accepted that Kripke in his *Naming and Necessity* (1980) demolished the old notion that only a priori knowledge can be necessary. Most importantly, he argued that there are cases of necessary a posteriori knowledge. The best-known example of this is the sentence "Hesperus is Phosphorus". Both "Hesperus" and "Phosphorus" are proper names and as such pick out the same thing in every possible world where that thing exists. For "Hesperus", the thing is the evening star, and for "Phosphorus", the morning star. But we know that "Hesperus" and "Phosphorus" actually refer to the same thing, the planet Venus. So the sentence "Hesperus is Phosphorus" is necessarily true: in all possible worlds "Hesperus" and "Phosphorus" refer to the same thing. Since the fact that both refer to the planet Venus was discovered empirically, Kripke argues, "Hesperus is Phosphorus" is a case of necessary a posteriori knowledge.

Could we use Kripke's strategy above to save the necessity of arithmetical knowledge in the contextual a priori account? At first sight, the argument looks promising, since the concept of necessity is expanded beyond strictly a priori truths—which arithmetical truths are not, according to the present account. The key point in Kripke's argument, however, is by no means guaranteed to apply to the case at hand. Kripke introduced the term "rigid designator" for terms that pick the same object in all possible worlds in which that object exists. In this he relied on a causal theory of reference. We can know that the term "Hesperus" picks out the evening star in every possible world where the evening star exists because there is causal link from the name "Hesperus" to the event of giving it its reference, the planet Venus.

But in the case of arithmetical knowledge, is there a causal link between, say, the numeral "two" and its reference? Can we think of numerals as rigid designators? Kripke himself was notoriously reluctant to discuss the finer details (e.g., the nonexistence of objects) of rigid designation. Among the distinctions he explicitly ignored was that of *de jure* rigidity and de facto rigidity (Kripke 1980, p. 21 n. 21). *De jure* rigidity is designation by stipulation, like in the case of Hesperus and Phosphorus, whereas de facto rigidity can be designation by description. In Kripke's example, "the smallest prime" is a de facto rigid designator because it picks out a unique object, the number two.

De facto rigid designation would seem to open up interesting possibilities. Let us continue with the contextual a priori account and deny that arithmetical knowledge comes from some kind of direct access to mind-independent natural numbers. In that case, whether or not there are such things as natural numbers, there could hardly be *de* *jure* designation for numerals. However, it appears that also in this case descriptions like "the smallest prime" *do* designate rigidly. Whenever we *do* have natural numbers, "the smallest prime" picks out the same number.

When it comes to natural numbers, the question of existence is obviously more tricky than with the planet Venus. The whole point of approaches like Pantsar (2014) is that we do not need to evoke mind-independent existence for numbers in order to have feasible epistemology of arithmetic. But at the same time, the theory is based on biological primitives and as such entails a high level of objectivity for arithmetical truths. There may not be such an object as the natural number two, but there is extremely strong objectivity in the way we use the numeral "two".[10]

However, it seems that the latter type of objectivity is enough for a Kripke-type theory of reference. "$2 + 1 = 3$" is true in all possible worlds where enough arithmetical or proto-arithmetical ability has developed to give reference for the numerals "2", "1" and "3". It is not necessary that there is a single object, the natural number two, to which we refer to when we use the numeral "two". Rather, "two" can designate different things at different levels. When a child first learns to use the numeral "two" correctly, her understanding of arithmetic is very limited and the numeral refers to her proto-arithmetical knowledge arising from seeing (or hearing or touching) two things.

When arithmetic develops beyond this primitive level, we find explicit ways of characterizing what "two" refers to, for example, by using the successor operation twice on zero. When we refer to a number like eleven, in which case our proto-arithmetical ability no longer suffices, we refer to using the successor operator on zero eleven times. That is how we learn to count and do basic arithmetic as children. When we use the numeral "two" in this more developed context, we are referring to something conceptually different from what the child does with her proto-arithmetical use of "two".

At a first glance, it would then seem obvious then that "two" cannot be a rigid designator in either the *de jure* or the *de facto* sense, as it can refer to different things. However, this is not so simple. It is clear that if we do not include arithmetical objects in our ontology, in the strict sense numerals cannot be rigid designators. But while in the contextual a priori account numbers are not postulated to exist, numerals are still thought to refer to something strongly objective, i.e., the concepts developed in arithmetic which are constrained by our biological structure. There is no single object that the numeral "two" designates, but it still manages to pick out something that is at the basis of animal and infant behaviour, as well as of our developed arithmetic. If arithmetic is constrained by biological primitives in a strong way as argued in the contextual a priori account, the numeral "two" picks out that something in all possible worlds where sufficiently developed biological organisms develop.

[10]It should be noted that Kripke also introduces (1980, p. 28) the term "strong rigid designation" for terms referring to necessary existents, presumably including numbers. It is unclear to me whether there are conceivable cases of strong rigid designation, but it should be clear by now that natural numbers should not be considered to be such.

This is not quite the rigid designation of Kripke, but it is essentially similar—as long as we turn the focus on cognitive abilities instead of objects. Rather than talk about all the worlds where the object exists, we focus on all the worlds where sufficiently developed biological organisms develop. Rather than talk about a single object, we focus on an ability with quantities that has a continuous development from the proto-arithmetical ability to developed formal arithmetic. With these adjustments, the use of the numeral "two" has essentially the same characteristic as Kripke's rigid designation: namely, it picks out the same thing in all the possible worlds where that thing exists.

What that "thing" is in case of arithmetic, however, is more complicated than in the case of Venus. At the stage of our development when we understand abstract objects, we can conveniently refer to the abstract natural number two. That is why the numeral "two" *appears* to be a rigid designator. However, under further analysis it is revealed that "two" refers to something much more complicated, a development which has different stages starting from the proto-arithmetical origins and ending up with formal axiomatic arithmetic. But importantly, that development is constrained by our cognitive architecture and thus guarantees a strong form of objectivity for the reference of "two". While this is not Kripke's rigid designation, it is rigid enough for our current purposes—explaining the apparent necessity of arithmetical knowledge.

Whatever the exact philosophical characteristics of natural numbers may be, our shared biological structure and the arithmetical knowledge built on that will make sure that the numeral "eleven" will pick out the same "object" from what we at the furthest stage of the development describe as the abstract domain of arithmetic, the set of natural numbers. But there is no need to postulate that such a set or its elements actually exist. Perhaps they do, but we can also explain arithmetical knowledge and the reference of numerals without evoking such platonic properties.

There are, of course, possible worlds in which arithmetic did not develop. But in Kripke's theory it is enough that the name picks out the same thing in all the possible worlds where the thing exists. When we adjust our understanding of arithmetical objects as above, a Kripke-like theory fits well with my theory of contextual a priori epistemology of arithmetic. It is enough that a numeral picks out the same thing in all the possible worlds where sufficiently developed biological organisms develop.

Kripke argued that knowledge can be necessary and a posteriori. If that is the case, it follows that arithmetical knowledge can be both necessary and contextually a priori. I hope to have shown that in a modified Kripkean framework we can explain at least the *apparent* necessity of arithmetical truths. Perhaps arithmetical truths are necessary also in the strong sense of Kripke, i.e. true in all possible worlds. The contextual a priori account does not take a stand on that question. But it can explain why arithmetical truths are true in all possible worlds with sufficiently developed biological organisms. For the purposes of this paper, that is enough. We have explained, based on the contextual a priori account of the epistemology of arithmetic, the last of the three apparent characteristics that arithmetical knowledge has. In addition to appearing to be both a priori and objective, it has now been shown why arithmetical knowledge appears to be necessary.[11]

References

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge: MIT Press.

Carey, S., & Sarnecka, B. W. (2006). The development of human conceptual representations. In M. Johnson & Y. Munakata (Eds.), *Processes of change in brain and cognitive development: Attention and performance* (Vol. XXI, pp. 473–496). Oxford: Oxford University Press.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics* (2nd ed.). New York: Oxford University Press.

Dehaene, S., & Brannon, E. (Eds.). (2011). *Space, time and number in the brain*. London: Academic.

Field, H. (1980). *Science without numbers: A defense of nominalism*. Princeton: University Press.

Frege, G. (1879). *Begriffsschrift*. Halle.

Frege, G. (1884). *The foundations of arithmetic* (J. L. Austin, Trans.). Oxford: Basil Blackwell, 1974.

Jenkins, C. (2008). *Grounding concepts*. Oxford: Oxford University Press.

Kant, I. (1781). *Critique of pure reason* (P. Guyer & A. Wood, Trans./Ed.). Cambridge: Cambridge University Press, 1997.

Kitcher, P. (1983). *The nature of mathematical knowledge*. New York: Oxford University Press.

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

Kripke, S. (1980). *Naming and necessity*. Cambridge: Harvard University Press.

Kuhn, T. (1993). Afterwords. In P. Horwich (Ed.), *World changes* (pp. 331–332). Cambridge: MIT Press.

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from*. New York: Basic Books.

Lewis, D. (1970). General semantics. *Synthese*, *22*, 18–67.

Maddy, P. (2014). A second philosophy of arithmetic. *The Review of Symbolic Logic*, *7*, 222–249.

Nieder, A. (2011). The neural code for number. In S. Dehaene & E. Brannon (Eds.), *Space, time and number in the brain* (pp. 107–22). London: Academic.

Pantsar, M. (2014). An empirically feasible approach to the epistemology of arithmetic. *Synthese*, *191*, 4201–4229.

Putnam, H. (1976). 'Two Dogmas' revisited. In *Realism and reason: Philosophical papers* (Vol. 3, pp. 87–97). Cambridge: University Press, 1983.

Spelke, E. (2011). Natural number and natural geometry. In S. Dehaene & E. M. Brannon (Eds.), *Space, time and number in the brain* (pp. 287–318). London: Academic Press.

Wittgenstein, L. (1956). In G. H. von Wright, R. Rhees, & G. E. M. Anscombe (Eds.), *Remarks on the foundations of mathematics*, Revised Edition (G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell, 1978.

Wright, C. (1983). *Frege's conception of numbers as objects*. Aberdeen: Aberdeen University Press.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–751.

Yablo, S. (2002), Abstract objects: A case study. *Noûs*, *36*, 220–240.

[11] This research was funded by the Academy of Finland, whose support is acknowledged with great gratitude. A big thank you is in place to the participants of the FilMat conference Philosophy of mathematics: objectivity, cognition, and proof in Milan. The discussions there helped formulate the arguments of this paper. The final version of this paper was written during a visit to the University of California, Irvine. I am grateful for the discussions with the staff there, especially with Sean Walsh, Penelope Maddy and Kai Wehmeier. Finally, with gratitude I note that the paper as it appears here benefited greatly from the thorough referee reports of two anonymous referees.

Chapter 6
Epistemology, Ontology and Application in Pincock's Account

Marina Imocrante

Abstract I submit that in Pincock's (Mathematics and scientific representation, 2012) structural account [SA] the request of a priori justifiability of mathematical beliefs [AP] follows from the adoption of semantic realism for mathematical statements [SR] combined with a form of internalism about mathematical concepts [INTmc]. The resulting framework seems to clash with Pincock's proposal of an "extension-based" epistemology for pure mathematics [EBE], in that the endorsement of [EBE] seems to ask for a form of conceptual externalism [EXTmc] that would not provide us with the a priori justifications for mathematical beliefs requested. I claim that Pincock's overall account of pure and applied mathematics would be made more stable if the assumption of [INTmc] was replaced by [EXTmc]. In eliminating [INTmc], [SA] would not entail any necessary commitment to forms of a priori justification for mathematical beliefs anymore, preventing the tension with [EBE]. Someone could object that the combination of [SR] and [EXTmc] would lead to ontological realism for mathematical objects [OR]. I answer by arguing that the kind of [EXTmc] that can be endorsed within Pincock's framework takes the content of mathematical concepts to be determined by contingent facts in the historical development of mathematical practice, so that no commitment to the existence of mathematical objects is required.

Keywords Pincock's epistemology of mathematics · A priori justification · Conceptual internalism/externalism

M. Imocrante (✉)
Vita-Salute San Raffaele University, Milan, Italy
e-mail: marina.imocrante@gmail.com

M. Imocrante
IUSS, Pavia, Italy

M. Imocrante
IHPST - Paris 1 Panthéon-Sorbonne University, Paris, France

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_6

6.1 Introduction

In his structural account[1] of applied mathematics [SA][2] (Pincock 2012) envisages a connection between the applications of mathematics in scientific representations and the necessity of a priori justifiability of the beliefs expressed by mathematical statements. At the same time, Pincock proposes an epistemological approach to pure mathematics that accounts for the content of mathematical concepts as closely tied and influenced by their development through the history of mathematical practice. After rehearsing some of Pincock's relevant assumptions, i.e. semantic realism for mathematical statements [SR] and semantic internalism for mathematical concepts [INTmc], I show that the stated need for a priori justification for mathematical beliefs [AP] follows from their combined adoption.

Let me first give a general sketch of the arguments to follow. After summarizing Pincock's proposal of an "extension-based" (Pincock 2012, p. 295) epistemology for pure mathematics [EBE], I argue that the [AP] request is questionable, in that it seems to clash with [EBE]. The reason is that in this framework the content of mathematical concepts seems to depend on their development through the history of mathematics; the justification of the beliefs expressed by mathematical statements thus ought to come from the knowledge of this history—this means that they would not be a priori justified. Pincock's move of appealing to a purely mathematical version of inference to the best explanation, in order to maintain the possibility of a priori justification of mathematical beliefs in his extension-based account, is discussed and rejected as implausible, by arguing that, given the way in which [EBE] is presented, there seems to be a tension between the latter and (semantic internalism and) the [AP] request.

To solve this tension, I claim that Pincock's accounts of both pure and applied mathematics would be made more stable if the assumption of [INTmc] were dropped in favor of what is taken to be the rival view on concepts, semantic externalism [EXTmc]. I argue that this replacement is licensed by the conception of mathematical concepts stemming from [EBE], and by its consistency with the treatment of mathematical concepts in [SA]. Given the necessity for [EBE] and [SA] to be consistent with each other, as parts of the same overall epistemological account, [INTmc] is rejected as questionable within [EBE] and replaced by [EXTmc], which is showed to be a legitimate semantic view about concepts which better coheres with both [SA] and [EBE]. Once the need for [EXTmc] within Pincock's account is advanced, I show that the joint adoption of [SR] and [EXTmc] does not entail any necessary commitment to forms of a priori justification of mathematical beliefs. The tension between Pincock's assumption of [AP] and his [EBE] proposal is thus avoided by eliminating the need for this assumption for the structural account of applied mathematics.

In conclusion, I answer to a possible objection to the adoption of a form of semantic externalism within Pincock's account. It could be claimed that the combination of

[1] Pincock calls his approach "structuralist account" (Pincock 2012, p. 27). I prefer to refer to it as the "structural account" not to confuse it with the positions we call 'structuralists' in the classical philosophical debate about mathematics.

[2] All the abbreviations in square brackets are mine.

[SR] and [EXTmc] leads to the endorsement of ontological realism for mathematical objects [OR]. This conclusion would be in contrast not only with the consistency, claimed by Pincock, of his account of applied mathematics with anti-realist positions, but also with the features of his epistemological account of pure mathematics. In reply to this objection, I emphasize that, given [EBE], the kind of semantic externalism which can be endorsed within his framework would not mandate taking the content of mathematical terms to be fixed by the reference to sui generis mathematical objects. Their content could be rather determined, in some way to be further explored, by contingent facts in the historical development of mathematical practice, so that no commitment to the existence of mathematical objects is required.[3] The history of mathematics is here seen as an experiential process progressing along developments of mathematical practice. The resulting view, offered as a possible elaboration of Pincock's extension-based proposal, has remarkable connections not only with Mark Wilson's account of concepts (Wilson 2006), as acknowledged by Pincock, but also with a form of naturalism for pure mathematics as the one endorsed by Penelope Maddy (Maddy 2011).

6.2 The Structural Account

Pincock's account of the different kinds of applications of mathematics to science [SA] envisages a structural relation between a physical concrete system and a mathematical structure. The relation between the two systems obtains "solely in virtue of the formal network of the relations" (Pincock 2012, p. 27) among the elements of the systems, e.g. some kind of morphism. Thanks to the existence of this structural relation, the physical system, on which the relevant properties are defined, is taken to instantiate the mathematical structure (Pincock 2012, p. 29). By means of this structural correspondence, scientific representations could be said to have a mathematical content. This content is what gives us the possibility of confirming our representations, by providing "conditions under which the representation is accurate" (Pincock 2012, pp. 25–26).

Pincock refers to the basic statements of a scientific representation as its "constitutive framework" (Pincock 2012, p. 121). Indeed, the different epistemic roles of mathematics in science form what Pincock calls *derivative representations*, and each one of them is taken to be grounded in a mathematical *constitutive framework*. On this view, mathematics provides the conceptual skeleton of scientific representations; the constitutive framework of a given scientific representation is taken to be its "foundational framework" (Baron 2013, p. 1168), composed by the basic scientific statements of the representation in which an essential use of

[3] This does not mean that a realist ontological position would be incompatible with this framework, but only that it is not implicated by Pincock's account.

mathematical terms is made. The acknowledgment of these statements makes possible the formulation and the comprehension of the scientific representation at issue.[4]

6.2.1 Semantic Realism

Central to the structural account is the assumption of semantic realism for mathematical statements [SR], in virtue of which we can consider them to be true—so that the relations of structural correspondence between mathematical structures and physical systems could be taken to hold. Indeed, Pincock starts by assuming mathematical platonism: "[...] the subject matter of mathematics is a domain of abstract objects [...] I take for granted that mathematical language is about these abstract objects" (Pincock 2012, p. 26). Later in the book, after a criticism of the indispensability argument for platonism (Pincock 2012, pp. 190–202), he weakens his assumption to realism about the truth-values of mathematical statements (Pincock 2012, p. 197). Pincock claims that "for science as it is practiced now, mathematical claims plays a central epistemic role" (Pincock 2012, p. 199). Given that mathematical statements can be said to play an "indispensable" role in science, where a statement is indispensable to a theory when "all competitors that remove that claim do worse by ordinary standards of scientific theory choice" (Pincock 2012, p. 202), he concludes that "we ought rationally to believe in the truth of some mathematical claims[5]" (Pincock 2012, p. 202). Semantic realism does not imply platonism, so that [SA] is compatible with both realist and anti-realist views about the existence of mathematical objects—once they agree in accepting mathematical statements as true (Pincock 2012, p. 197).

6.2.2 Internalism About Reference and Conceptual Internalism

Given that [SA] requires nothing more than realism about the truth-values of mathematical statements to account for their role in scientific representations, it would seem plausible to ascribe to [SA] a form of what Linnebo (2012) calls "semantic reductionism". Linnebo maintains that we are endorsing a semantic reductionist perspective when, from the entitlement to adopt the language of a given domain (e.g.

---

[4]For example, the scientific content of the law of universal gravitation is made possible by the constitutive framework provided by Newton's laws of motion, in which mathematical terms constitutively appear - unless, of course, one agrees with Field (1980).

[5]It thus seems fair to say that Pincock could in principle accept a form of indispensability argument concerning the truth of mathematical statements (i.e. an argument of the form of the one stated in Pincock (2012), p. 202). However, as we will see when discussing the a priori justifiability request for mathematical beliefs, he does not think that an a posteriori argument can give us conclusive reasons to believe mathematical claims to be true. For Pincock, in order to give to mathematical statements the central role they have in our scientific representations, we need to have reason to believe them to be true prior to their use in science.

mathematics) for some kind of objects (e.g. mathematical objects), we do not conclude to be justified in attributing to such language reference to the relevant objects, in that if "the principles governing the use of [the language at issue] make no mention of abstract objects [...]" then "to ascribe reference to such objects to terms of [the language at issue] seems to be to go beyond what is warranted by the linguistic data. [...] On the reductionist interpretation a formula is true relative to a string of objects, just in case the formula is assertable relative to that string" (Linnebo 2012, p. 52).

In fact, following [SA], concerning mathematical structures and their objects the only thing that seems to be requested is to be able to consider mathematical assertions true—in order to be entitled to institute some kind of structural relation between mathematical structures and physical systems. Every further non-reductive semantic interpretation of mathematical statements would seem to be unnecessary for [SA] to apply. This could lead to adopt a radical form of semantic inferentialism for mathematical terms, for which the semantic content of a term would be determined by the rules governing inferences to and from it Brandom (1994) and the notion of reference would not play any significant role. Nevertheless, Pincock refuses to deny that mathematical terms have reference[6]: he takes [SA] to allow that, if we are entitled to adopt mathematical language, taking mathematical statements constitutively used in our scientific representations to be true, then we can be justified in attributing to the terms of such language reference to mathematical objects. This position corresponds to the assumption of a form of internalism about reference, to be distinguished from semantic inferentialism.

Along with this view, a form of semantic internalism for mathematical concepts [INTmc] is assumed by Pincock. The form of conceptual internalism he endorses is quite peculiar[7]: quoting Pincock, "[...] I assume that our mathematical concepts contribute all the mathematical content that a representation may have. Among other things, this involves the semantic internalist assumption that the concepts are sufficient to pick out their referents" (Pincock 2012, p. 26). In this framework, the content of a concept exclusively ought to derive from the features of the concept itself. It is difficult to learn more about Pincock's view of semantic internalism, because his explanation stops here. As we saw, Pincock does not deny that mathematical terms have reference, so that his adoption of a form of conceptual semantic internalism does not imply a form of radical semantic inferentialism for mathematical terms; he rather denies that external world's aspects could determine their reference, claiming that only the concept's definition can determine it. On the contrary, semantic externalism is assumed for physical concepts, since "[...] features of the world beyond what the subject is aware of could be part of the content of [scientific] representations" (Pincock 2012, pp. 26–27).

[6]Pincock even seems to reject internalism for mathematical terms, when, speaking about the rules for the use of mathematical terms, he claims that "a plausible requirement is that these rules must correspond to the genuine features of the things I wind up referring to" using those terms (Pincock 2012, p. 126).

[7]See Goldberg (2007) for a review of the philosophical debate between internalism and externalism in epistemology.

6.3 A Priori Justifiability

Within the structural account, the basic statements of a scientific representation (i.e. the scientific statements making an essential use of mathematical terms that form the constitutive framework of the representation at issue) ought to be a priori justified. Pincock appeals to a notion of a priori as the one endorsed by Albert Casullo (2003): a given belief is 'a priori justified' if it is "[non experientially justified", while this does not entail that it cannot be defeated by empirical evidence (Casullo 2003, p. 38). A defeasible belief is a belief whose justification is fallible, i.e. a belief that could be invalidated when further evidence is available (e.g. experiential evidence).[8] In this way, he can account for the possibility that some constitutive representations eventually come to be disconfirmed through scientific evidence in the history of scientific practice, and the beliefs in their statements withdrawn despite the a priori character of their initial justification. To justify a priori a given mathematical belief means for Pincock to justify it (i) independently from empirical evidence (Pincock 2012, p. 286), (ii) internally to pure mathematics (Pincock 2012, p. 295) and (iii) prior to its use in science (Pincock 2012, p. 286). What Pincock has in mind is an epistemic priority: if we do not have reason to believe the relevant mathematical statements to be true, then we will not have reason to believe the scientific representations that involve those statements to be true.

6.3.1 (Absolutely) A Priori Justifiability of Mathematical Beliefs

On the other hand, concerning purely mathematical beliefs, Pincock explicitly states the need to have an *absolutely* a priori justification for those which are employed in the constitutive frameworks of scientific representations (Pincock 2012, p. 140). The absolutely a priori character of the justification of these mathematical beliefs entails that they have to be indefeasible. Quoting Putnam, they become "truths which it is always rational to believe, nay, more, truths which it is never rational to even begin to doubt" (Putnam 1983, p. 90 in Casullo (2003), p. 28). The a priori justifiability of mathematical beliefs [AP] is needed to give us reason to believe our pure mathematics to be true, and consequently it is what allows us to use mathematics in building and confirming our scientific representations. The practical demand of providing a justification for mathematical beliefs which is epistemically prior to their use in science is capital in [SA]: Pincock refuses forms of indispensability argument as inappropriate to justify (a posteriori) our beliefs on the truth of mathematical claims, because "if this understanding [of the scientific process of representation, prediction and testing] really presupposes that the agent already believes that the mathematical claims are true, then this process cannot be the main source of justification for these

[8] In this respect, Pincock refers to [SA]'s constitutive frameworks as both "a priori and relative" (Pincock 2012, p. 138).

beliefs" (Pincock 2012, p. 217). The truth of mathematical claims constitutively used in sciences must be established before their use in scientific representations, and does not depend on the scientific success of scientific representations employing those claims; as a consequence, their truth is not to be rejected in case of failure of the scientific representations at issue.

6.3.2 Where the Need for [AP] Comes from

If we take a closer look at Pincock's structural account of applied mathematics, we can track the origin of the stated need for a priori justification of mathematical beliefs [AP]. We saw that Pincock adopts [SR], and he combines it with [INTmc]. He insists on the reasons agents must have to believe in the scientific representations they accept: semantic realism for mathematical statements is the precondition for believing the scientific representations in which they occur to be true. At the same time, endorsing the internalist assumption, the concepts expressed by mathematical statements ought to provide all the mathematical content of scientific representations. This means that mathematical statements are not allowed to be true in virtue of some external referent—they are true in virtue of their own content, so that the possibility of finding some external justification for mathematical beliefs is automatically ruled out. The endorsement of [INTmc] in combination with [SR] generates the need for justifying the beliefs expressed by true mathematical statements internally to pure mathematics and independently from any other evidence. Moreover, Pincock's [SA] requires them to be justified prior to their use in scientific representations, in that agents must have reasons to hold the relevant mathematical statements before they are applied in science. Notice that these three conditions—internally to pure mathematics, independently from any other evidence and prior to the use in science—correspond to the criteria (i), (ii) and (iii) identified above as defining the (absolutely) a priori character of the justification of mathematical beliefs. This is why within [SA], where mathematical beliefs are to be justified prior to their use in science, the combination of [SR] plus [INTmc] generates the need for their justification to be (absolutely) a priori.

6.4 Extension-Based Epistemology

Pincock's epistemological project for pure mathematics, namely his "extension-based epistemology" [EBE] (Pincock 2012, p. 296), is inspired by two requirements: justifying mathematical beliefs a priori on the one hand, and taking into account the historical development of mathematical concepts on the other.

Indeed, the main goal of Pincock's account is to show how a plausible epistemology of mathematics should be backed by appropriate reconstructions of the history of mathematics. His project aims at giving voice to the naturalistic requirement that the reconstruction of the development of mathematical concepts and theories through the history of mathematical practice should ground the epistemology of mathematics (Pincock 2012, p. 299). Pincock proposes his own view of what he calls "historical extensions" of mathematical concepts (Pincock 2012, p. 293) for which he draws on the account of physical concepts developed in Wilson (2006). According to Wilson, the complexity of concepts can be tracked by thinking of several networks of complementary areas of significance (Wilson calls those network "facades" (Wilson 2006, p. 7), while their areas of significance are named "patches" (Wilson 2006, p. 41)). This picture is meant to account for the "multi-valuedness" of our concepts (Wilson 2006, p. 456). Unambiguous categorization of objects can still be made depending on the practical goals at issue: subjects may classify a given object in a particular area of significance, and later discover its affinity and relations with other areas, in a gradual process of extension of their knowledge. To explain his view of the complex conceptual networks we deal with, Wilson uses the metaphor of an atlas, in which each map represents the geography of the earth maximizing some representational features and overlooking others. To have a full understanding of the geography of the earth, we need to combine the information supplied in several maps, in a way depending on our practical goals (Wilson 2006, pp. 289–294). Wilson insists that our concepts are not to be conceived as immutable entities, for they constantly evolve depending on contexts and practical purposes.

Pincock evokes a similar process of conceptual extension for mathematical concepts. The term 'extension'[9] here refers to the possibility for a given mathematical concept to evolve and modify its features through the historical development of mathematics, generally by expanding its range of application. In particular, he focuses on the example of the extension of the concept of square root function (Pincock 2012, pp. 268–275): despite its original definition as a function on non-negative real numbers, the square root function came to be defined for complex numbers. This extended definition introduced some remarkable changes in the function?s behavior and the optimal setting to study it turned out to be a Riemann surface. The example is again borrowed from Wilson (Wilson 2006, pp. 312–319), who compares the evolution of physical concepts to the extension of real-valued functions on the complex plane, where they behave differently as they display multi-valuedness. Pincock elaborates on this example in more details. The evolution of the concept of square root function from the real to the complex plane is taken to be the prototype of the kind of mathematical extension that Pincock has in mind:

> [...] With the square root function, [...] what an extension-based epistemology must confront are the steps that mathematicians like Riemann and Weyl took to argue that, in fact, the best way to approach this function is a [...] function on the Riemann surface. This argument is ampliative [...]. It involves a conclusion that goes beyond previously known theorems and the relevant concepts. (Pincock 2012, p. 296)

[9]Obviously we are not dealing with the traditional linguistic nor logical conception of 'extension' of a concept.

6.4.1 Looking for a Purely Mathematical Form of Inference to the Best Explanation?

According to Pincock, the possibility of justifying a priori mathematical beliefs may be grounded in a "purely mathematical version of inference to the best explanation" (Pincock 2012, p. 296). With [EBE], he aims to account for the fact that, throughout the course of the history of mathematics, mathematicians have made discoveries that seem to involve a kind of ampliative reasoning, rather than just depend on direct proofs from axioms; this sort of conceptual extension has taken mathematicians beyond known beliefs, as in the case of the square root function.[10] While the learning of the historical developments of mathematical concepts and theories would play a role in the starting of our mathematical knowledge, it would not be central to the justification requested for mathematical beliefs; in contrast, what provides evidence for the extensions of mathematical knowledge would be explanatory considerations, conceived as internal to mathematics. In the case of the concept of square root function, the appeal to the representation on a Riemann surface would constitute an help to the explanation of the peculiar multi-valuedness of the function at issue, like forms of inference to the best explanation do in cases of scientific explanation. In order to account for the possibility of justifying mathematical beliefs through a purely mathematical inference to the best explanation, Pincock claims:

> With my approach [...] agents must obtain evidence for the existence of the extensions they propose. This evidence may come from reflection on properly grounded concepts, or it may come from proofs whose premises are not all based on the features of the relevant concepts.[11] In particular, we allow explanatory considerations to provide evidence for mathematical claims. (Pincock 2012, p. 298)

Pincock thus accept that the growth of mathematical knowledge is an historical process of development, made by a progression of practices in mathematics, but he does not want this historical process itself to be the source of our evidence for the extensions of the content of mathematical concepts—the corresponding explanatory process occurring within mathematics would provide this evidence.

Following this line of thought, the assumption of the need for [AP] would thus be compatible with Pincock's epistemological account of pure mathematics (Pincock 2012, p. 297): even if he claims that an adequate epistemology for mathematics must be capable of reconstructing historical developments in mathematics in a plausible way, he would not be committed to hold that mathematical beliefs must be justified via a posteriori knowledge about the history of mathematics. However, it is questionable

[10]In addition to the case of the square root function, Pincock's examples concern the development of group theory and theory of limits (Pincock 2012, p. 297), the arithmetical properties of addition (Pincock 2012, p. 298) and the acceptance of the set-theoretical framework (Pincock 2012, p. 298–299).

[11]This is why, regarding mathematical concepts, Pincock refuses an exclusively concept-based approach, as the one endorsed by Peacocke (2004), together with an opposite empiricist approach as the one proposed by Jenkins (2008).

if the appeal to a purely mathematical version of inference to the best explanation can safeguard the a priori character of Pincock's epistemology. First of all, this renders Pincock's overall account precarious, as Sam puts out:

> [...] Mathematics must be true, but its truth must be justifiable a priori. [...] Pincock makes some tentative suggestions toward such an account, the gist being that a purely mathematical version of inference to the best explanation may be able to ground the truth of mathematical claims a priori. But even this requires the assumption that some basic mathematical claims are already justified and so does not constitute a complete account. So the resulting view is somewhat precarious [...].? [Baron 2013, p. 1171]

Most important, consider that inference to the best explanation is an abductive argument where the conclusion stands as the hypothesis, selected among others, that if true would offer the best explanation of the facts described in the premises. In this case, the fact to be explained is a given mathematical beliefs, and the proper explanans is taken to be constituted by considerations internal to mathematics. Given that Pincock claims that we can have an a priori form of inference to the best explanation, those considerations ought to be entirely a priori, as well as the resulting mathematical belief would be a priori justified. But, with regard to the formulation of the explanatory considerations at issue, does the [EBE] account provide us with the resources to distinguish between a form of evidence purely intra-mathematical, a priori in the sense that it would be obtained by pure reasoning, and a form of evidence coming from our empirical experience, as the evidence provided by the knowledge of the historical process of introduction and evolution of new mathematical concepts would be? For example, in the case of the example of the square root function, it seems that an historical aspect (i.e. the discovery of Riemann surfaces) enters necessarily in the formulation of the explanation of our belief in the multi-valuedness of the function. Pincock acknowledges that the representation on a Riemann surface "helps explain" this multi-valuedness (Pincock 2012, p. 296) and he does not offer arguments to conclude that to say this would have to be considered in a different way from saying that an historical, external fact of the development of mathematics (i.e. the discovery of Riemann surfaces) contributed in a new essential way in the definition of the content of our concept of square root function. The distinction between a posteriori and a priori evidence—the historical process of development of mathematical notions on the one hand, and the explanatory process internal to mathematics on the other—is not easy to state following Pincock's sketch of the mathematical form of inference to the best explanation. In a passage of his book, Pincock himself points out that:

> It is not desirable or possible [...] o factor our representations into contents arising from conceptual resources alone and contents arising from the world independently of these conceptual resources.[12] Our concepts are necessary to get the process started because without them we would not be able to refer to purely mathematical structures or physical features of our target systems. But this minimal role in starting the process of mathematics and science

[12]This line of thought could be taken to be reminiscent of the Quinean claim of the impossibility of a clear-cut distinction between the analytic and the factual component of the truth of a statements (Quine 1951).

> is consistent with features of the mathematical domain and the physical world playing a more significant role in prompting new patches of successful representation. [...] Learning more about the world leads us to assert and test new claims beyond those licensed by our concepts. This flexibility is crucial to progress in mathematics and science. (Pincock 2012, p. 277)

Extending this reasoning, I submit that within [EBE] it remains difficult to distinguish between a priori and a posteriori evidence for the best intra-mathematical explanation of mathematical beliefs.[13] I thus doubt that the claim of the a priori justifiability of mathematical beliefs can rest upon the possibility of a purely a priori form of inference to the best explanation.

6.5 Tension Between the a Priori Justifiability Request and the Extension-Based Proposal

I take Pincock's epistemological approach to pure mathematics to chiefly rely on the historical episodes in the development of particular mathematical concepts; we know whether we are justified in believing our mathematical claims to be true by looking for a "world-driven story" (Pincock 2012, p. 276) for the development of mathematics. The method of inquiry that Pincock proposes goes in this direction: he presents the reader with a detailed example of the analysis of the historical development of a particular mathematical concept, and draws theoretical claims from it. I suggest that a plausible consequence of a similar position, drawing theoretical conclusions from historical examples, would be to see the facts of the historical development of mathematical practice as entering into the characterization of the content of mathematical concepts. Following the line of thought I propose, the content of mathematical concepts would be seen to derive from their progressive development through the history of mathematical practice: historical aspects, as the discovery of Riemann surfaces in Pincock's example, enter into the extended definition of the concept at issue—in the example, the one of square root function.[14]

If historical, external facts could enter into the definition of mathematical concepts, internalism no longer seems to be the appropriate semantic position regarding these concepts. Similarly, I propose to integrate Pincock's [EBE] by looking at the history

[13]It could be objected that the uncertainty about the evidence for the justification of mathematical beliefs (a priori or not) would not be a problem for Pincock's account, in that we did not show that the uncertainty occurs in every case, while we only suggested that it occurs in some case—i.e. in the case of the example. But notice that, from a philosophical point of view, the fact of not having the resources to distinguish between a priori and a posteriori evidence in some case means that the account at issue does not provide the resources to distinguish between the two kinds of evidence in every case—and this is the problem we underscore.

[14]Frege famously warned us to distinguish the history of a concept from its definition (Frege 1903, Sect. 56–67). Here I do not deny this distinction; I rather claim that in the context of Pincock's [EBE] we do not seem to be in the position to make this distinction, given that historical aspects seem to enter into the definition of the content of mathematical concepts.

of mathematical practice as a process of different experiences regarding mathematical practice that follow one another in time—while it is not conceivable as an ordinate system of a priori statements. If we adopt the view I propose, our knowledge about mathematics (the access to mathematical statements and their truth-values) would necessarily be a posteriori.

I take my view to be sympathetic with Pincock's [EBE] proposal, even though it would demand a conceptual shift from his stated [AP] request. From [EBE], we learn that we should consider the content of mathematical concepts as depending on the extensions they undergo through the historical development of mathematical theories. Consider again the already mentioned example of the extension of the concept of square root function: thanks to the progress made in the history of mathematics and to the discovery of new ways of representing that function, we enlarge our comprehension of this concept, "discovering features" that "we had not previously been aware of" (Pincock 2012, p. 274). In this paradigmatic case, the reasons we have to believe that the square root function is a multi-valued function, optimally represented on a Riemann surface, come from our knowledge of the gradual historical process of mathematical development occurred. Historical aspects enter essentially into our comprehension of the content of the concept of square root function: without the particular historical contingency of the introduction of the Riemann surfaces, we would not have been able to discover a feature of the concept at issue, i.e. its multi-valuedness. In general, this entails that the reasons a scientist has to believe some statements of pure mathematics to be true can come from specific historical cases of development of this and that mathematical concept. Facing this historical setting, we can consider whether the request of a priori justifiability of mathematical beliefs would not become problematic. Indeed, it is difficult to see how the justification of the beliefs expressed by mathematical statements can be a priori, given that historical considerations seem to count essentially in their favor. One should bear in mind that the first feature of a priori justification is independence from experience. On the contrary, the empirical events in mathematical practice occurring through the history of mathematics seem to count as fundamental ingredients in the justification of the beliefs expressed by statements of pure mathematics. Pincock is aware of this problem. When he comes to present a case study from pure mathematics concerning our entitlement to believe the theorems of group theory to be true, he claims:

> It should be clear in what sense this entitlement is a priori. [...] No perceptual experiences or perceptual beliefs figure in the justification of the claim that there is no element in any group that lacks an inverse. The justification begins with the axioms of group theory and involves a simple application of logical inference rules. No appeal is made in the proof to what the agent perceives. The hope, of course, is that the same sort of thing can ground our entitlement to the theorems of group theory even when their justification involves a long series of steps. (Pincock 2012, p. 289)

The analysis of the example's details would thus be one of the experiences necessary to acquire the mathematical concept at hand, and would not be part of the justification of the belief in the truth of the mathematical theorem at hand. Moreover, the absence of 'perceptual' elements in the justification should guarantee its a priori character. But it remains unclear how the mathematical knowledge obtained comes to be justified without making reference to the historical development described in the example. The distinction "between the experiences necessary to acquire a concept and the experiences (if any) that figure in the justification of a belief involving that concept" (Pincock 2012, p. 289) is in itself clear, but it is unclear how the distinction could be maintained within the history-driven account of [EBE] proposed: here the general necessity of taking into account the historical development of concepts and theories through the history of mathematics is not clearly separated from the understanding of particular cases of the development of certain concepts and theories.[15]

The experiences involved in the learning of the historical development of mathematical concepts and theories cannot count as a priori, although they may be not strictly speaking perceptual. A justification obtained in this way will be internal to mathematics (ii) and can precede the possible use in science of the corresponding mathematical beliefs (iii), but it will not be independent from experience in the way required by the criterion (i) of Pincock's own definition of a priori justification. In the same way, what emerges from Pincock's main example is that our experience of the historical development of the concept of square root function essentially figures in the justification of our new mathematical beliefs about the function. The considerations internal to mathematics that Pincock offers to justify mathematical beliefs ultimately consist in the learning of the historical development of the concepts and theories at issue,[16] and such a process of learning cannot be conceived as part of an a priori justification of mathematical beliefs.

The demand of a priori justifiability of mathematical beliefs thus seems to clash with my reading of Pincock?s extension-based epistemology.[17] I suggest that such a request should be discarded, in order to favor a conception of the justification of mathematical beliefs more fine-tuned with [EBE].

6.6 Eliminating the Tension

I argued for the unlikelihood of Pincock's appeal to a purely mathematical form of IBE as a way of solving the tension between [AP] and [EBE]. I suggest that the problem could be avoided by renouncing to the a priori justifiability request.

First of all, it seems plausible to assume that [SA] and [EBE] ought to be consistent to each other, in that they are presented by Pincock as parts of the same overall epistemological account of mathematics (applied and pure). Now, while in describ-

[15]The description of Pincock's epistemological proposal for pure mathematics is difficult to articulate leaving behind the specific features of the examples proposed; this suggests that in this account some prior experience and understanding of particular cases seems to count in the justification of beliefs involving the mathematical concepts at issue.

[16]Remember we think about the history of mathematics as an experiential process within mathematical practice.

[17]Pincock himself, in a previous work, claimed that a conception of concepts built on Wilson's open-ended account would abstain from a priori considerations (Pincock 2010, p. 116).

ing [SA] Pincock explicitly assumes a form of conceptual internalism [INTmc], my analysis of his [EBE] rather suggests the possibility of endorsing an externalist view about mathematical concepts [EXTmc]. If [SA] and [EBE] are supposed to be coherent parts of the same overall epistemology of mathematics, the endorsement of a different view about concepts in the two accounts should not be allowed. It thus seems that we will have to choose either [INTmc] or [EXTmc], in order to give a coherent picture of Pincock's overall account.

6.6.1 Substituting Conceptual Internalism with Externalism Within Pincock's Account

In choosing between [INTmc] and [EXTmc], one should immediately notice that the internalist assumption does not seem to be coherent with the open-ended view on mathematical concepts that [EBE] promotes. In recalling that within [EBE] historical, contingent aspects are allowed to enter into the characterization of the content of mathematical concepts, we see that this position is perfectly coherent with a rejection of an internalist view on mathematical concepts (and with refusing to adopt different semantic approaches for mathematical concepts and physical ones, as initially stated). It thus seems natural to suggest that when he comes to articulate his proposal for the epistemology of pure mathematics, Pincock should reject [INTmc], in that it is not compatible with the open-ended view on mathematical concepts adopted in [EBE].

On the other hand, it seems to me that the opposite view on concepts, i.e. semantic externalism, could be the most suitable semantic view about concepts for Pincock's extension-based proposal for pure mathematics. Remember that in [EBE] the historical (external) facts of the development of mathematical practice are taken to count as ingredients in the justification of the beliefs expressed by statements of pure mathematics, and in the definition of mathematical concepts. Moreover, contrary to Pincock's internalist assumption, a form of semantic externalism could fit well with his structural account of applied mathematics too. In fact, as we saw, Pincock never denies that mathematical terms have reference, and together with the endorsement of semantic realism for mathematical statements he refuses forms of semantic reductionism and inferentialism: in [SA] as presented by Pincock, if we are entitled to adopt mathematical language, then we can be justified in attributing to the terms of such language reference to some kind of objects. This is why I do not see any objection to the possibility to ascribe an externalist view of mathematical concepts to the proponent of [SA]: following Pincock's structural account, we could take the content of mathematical concepts to be determined by external considerations, e.g. historical facts of the development of mathematical notions, and the reference of mathematical language to be reference to these objects (these historical facts).[18] On the contrary, I object to the attribution of an internalist view on concepts to the proponent of [EBE], given the historical treatment of mathematical concepts pursued by this account.

[18] Section 6.7.2 outlines an attempt to characterize this particular kind of externalism.

I thus suggest that the [INTmc] of Pincock's framework should be replaced by a form of semantic externalism. The stated form of internalism does not fit well with the open-ended view on mathematical concepts promoted by [EBE], while an externalist view [EXTmc] would fit well with both [SA] and [EBE]—this replacement is consistent with both the semantic treatment of mathematical concepts that we could make in Pincock's structural account of applications, and the understanding of mathematical concepts that comes out from his epistemology of pure mathematics.

6.6.2 Eliminating [AP]

By adopting an externalist approach to mathematical concepts, we can account for their open-ended features, as suggested by [EBE]: the content of mathematical concepts depends on their development through the history of mathematical practice, where they evolve thanks to the progress we make in our mathematical theories. Accordingly, the justification of our mathematical beliefs ought to be conceived as external, in that it comes from the knowledge of these historical developments, rather than being a priori. The need for the a priori justifiability of mathematical beliefs [AP], that we saw to stem from the combination of semantic realism and conceptual internalism, can thus be eliminated once that [INTmc] is replaced by [EXTmc]. Consequently, in eliminating [AP], the stated tension between this assumption and the [EBE] proposal for pure mathematics would be avoided.

6.7 A Possible Objection

I presented a viable path to replace the internalist view on mathematical concepts with an externalist one, staying within Pincock's account, so that we would no longer be committed to claim the a priori justifiability of mathematical beliefs. The tension with the extension-based proposal would thus be eliminated. However, there is a possible objection to the adoption of a form of semantic externalism for mathematical concepts within Pincock's account, that should be considered. Someone could claim that the endorsement of semantic externalism for mathematical concepts, together with semantic realism concerning mathematical statements, leads to accept a form of ontological realism for mathematical objects.

6.7.1 Are We Committing [SA]'s Proponent to a Form of Realism for Mathematical Objects?

Consider that a standard form of semantic externalism usually takes the content of concepts to be fixed by the reference of terms to objects, conceived as external and independent from the understanding of subjects. If we claim that our mathematical statements are true [SR] and that the meaning of their terms depends on external referents, whose features are independent from subjects, then it seems that we have to conclude that mathematical statements are true in virtue of the existence of their external referents, i.e. that there are objects to which these statements are about, whose existence is independent of us and our thought and practices. In an externalist perspective, we could think that only the existence of mathematical objects gives us reasons to claim the truth of our statements about them. It is easy to see that this thesis corresponds to a form of ontological realism for mathematical objects [OR]. So, if semantic realism and a standard form of externalism are endorsed, then it seems that the objects to which statements expressing the content of mathematical concepts make reference ought to be conceived as existing independently of our understanding. [OR] would follow, for this thesis exactly states that the objects which mathematical statements are about exist as fixed entities, independent of our understanding of them.

However, the endorsement of [OR] would be in contrast with Pincock's epistemology for both applied and pure mathematics. On the one hand, regarding [SA], we saw that this account is taken to be consistent with anti-realist positions. If it ended up to be committed to ontological realism for mathematical objects, then its openness to anti-realist positions sharing a structuralist setting would not be possible. On the other hand, [OR] would also clash with the open-ended character of mathematical concepts as stated by Pincock's extension-based epistemology for pure mathematics. As showed by the example of the square root function, in [EBE] mathematical concepts are not conceived as fixed entities, for they evolve through the development of mathematical theories and depending on the practical purposes of the representations at issue—e.g. the concept of square root function refers to a mathematical operation that maps a set of real numbers onto itself, but throughout the historical development of mathematics it comes to refers to a function on complex numbers that could be represented differently. If we look at mathematical concepts in this way, a conceptual evolution can imply a change in reference: the content of our concepts cannot be simply fixed by the reference to some objects independently of our understanding and practices, as an externalist adopting ontological realism for mathematical objects would maintain, in that its reference may change through the development of mathematical practice.

The conflict of ontological realism for mathematical objects with Pincock's epistemological account for applied and pure mathematics thus seems to be inescapable. To prevent this from happening, the conclusion that the combined adoption of [SR] with [EXTmc] commits to this kind of realist ontology for mathematics should be avoided.

6.7.2 *No, We Are Not: A World-Driven Story for Mathematical Concepts*

To answer this objection, we have to consider the specific form of conceptual externalism [EXTmc] that may be adopted within Pincock's account. As we saw, in the extension-based account, the content of mathematical concepts depends on their development through the history of mathematics. Coherently, [EXTmc] cannot take the content of concepts to be conclusively fixed by the reference to objects, conceived as unchanging entities. Their content seems rather to be determined by the contingent facts of the historical development of mathematics, external to subjects but depending on the progress of their mathematical practice and their understanding. Only a form of semantic externalism for mathematical concepts which accounts for these features can be consistent with Pincock's account.

To see what those *facts* of the historical development of mathematics can be about, consider for example Penelope Maddy's naturalistic conception of the "objective facts that underlie mathematical practice" (Maddy 2011, p. 112), as expressed in Maddy (2011). Her examples refer to concept formation in set theory and group theory and to the different formulations and applications of the Axiom of Choice (Maddy 2011, pp. 78–81). For example, the concept of group, originally used to study permutations and the solvability of algebraic quintic equations, went on to be recognized as the appropriate tool to study symmetry, and today group theory essentially occurs in model constructions within different scientific contexts. The variety of use of the concept of group, and its capacity to unify different structures which share several properties, represent what Maddy would call a mathematically 'deep' fact.[19]

Notice that Maddy's mathematical 'facts' should not be seen as theoretic facts, but they can be considered as the historical facts of the fruitful use of particular notions, statements and theories during the history of mathematical practice, in thinking about the history of mathematics as a gradual process. In this sense, her conception of pure mathematics is not far from the open-ended view proposed by Wilson and inherited by Pincock. Pincock's examples of the extensions of mathematical concepts seem to correspond to the occurrences of certain uses of mathematical notions which turn out to evolve (in fruitful ways) through the developments of mathematical theories during the history of mathematical practice. It seems to me that a form of naturalism for pure mathematics of the kind presented by Maddy is not far from Pincock's perspective, even if he does not explicitly refer to Maddy's position. This correlation could be useful to understand Pincock's semantic view on concepts as a form of semantic externalism that takes the content of our mathematical concepts to be determined by the historical facts of the evolution of mathematical practice: during the history of mathematics, while different mathematical theories arise, different changes in the relevant concepts follows one another.[20] The content of mathematical concepts could be seen as the result of this historical evolution.

In adopting this specific form of [EXTmc], together with semantic realism for mathematical statements, no commitment to the existence of mathematical objects would be requested. In fact, the content of mathematical concepts would not depend

[19]Maddy focused her inquiry on the phenomenon of what she labels "mathematical depth" (Maddy 2011, p. 112).

[20]In what I take to be a similar perspective, Shapiro talks about possible "changes in meaning" between the various mathematical theories (Shapiro 2014, pp. 320–325).

on the reference to some external object—it would rather depend on the changes they go through during their evolution through the history of mathematics. In this perspective, in endorsing [EBE] one would be committed to the existence of the (external) facts of the evolution of mathematical theories, but not to the existence of specific mathematical objects. Notice that these 'facts' are to be conceived as historical entities, not as fixed concrete objects: this would offer to the proponent of a similar position several ways to elude the classical objection raised by Benacerraf (1973) for views taking mathematical concepts to refer to concrete physical objects strictly speaking. Obviously, while here I am only outlining a proposal, a form of semantic externalism for mathematical concepts based on the facts of mathematical practice is extremely difficult to devise. It would be a massive undertaking that would require further work to be developed on entirely new basis.

6.8 Conclusion

We saw that, in Pincock's epistemology of mathematics, [AP] follows from the adoption of [SR] combined with [INTmc], and that the resulting framework seems to clash with the [EBE] proposal for pure mathematics. Given the problems posed by the request of a priori justification of mathematical beliefs for the epistemology of pure mathematics proposed, I suggested to replace [INTmc] with [EXTmc], showing why this replacement could be legitimate in staying within Pincock's epistemological account for both applied and pure mathematics. Once the assumption of [INTmc] is discharged, no necessary commitment to a form of a priori justification of mathematical beliefs follows. In eliminating the need for the assumption of [AP] from [SA], its tension with [EBE] is removed. Finally, a possible objection to the adoption of a form of semantic externalism in Pincock's account is raised and answered by describing the specific form of [EXTmc] that may be endorsed within this account.

In conclusion, I would like to focus the attention on the view of mathematical concepts emerging from Pincock's work. Consider that the capacity of combining the analysis of different case studies in a multifaceted and versatile epistemological account can be taken to be the main virtue of Pincock's epistemological account of applied mathematics. At the same time, a similar goal can be seen to animate his inquiry in the epistemology of pure mathematics: his way of building on certain examples of the evolution of particular mathematical concepts, describing their "historical extensions" (Pincock 2012, p. 293), proves it. From this point of view, we can see that the attention to the practice and the history of mathematics is central to Pincock's account. This attention should be maintained when we come to describe the features of mathematical concepts. As we saw, Pincock himself declares to endorse Wilson's view on concepts, seeing mathematical concepts as open-ended and subject to evolution (Pincock 2012, p. 265). These features should not be forgotten when there is to take into account the specific form of semantic view about concepts that can be consistent with Pincock's epistemological account of mathematics.

Acknowledgments I would like to thank Francesca Boccuni, Marco Panza, Andrea Sereni, the audience of the *First International Conference of the Italian Network for the Philosophy of Mathematics* (Milan 2014), the participants of the seminar *Séminaire Inter-Universitaire de Philosophie des Mathématiques de Paris 1 et Paris 4* (Paris 2015), and two anonymous referees for their helpful remarks that led me to significantly improve this paper.

References

Baron, S. (2013). Book review Mathematics and scientific representations. *Mind*, *122*(488), 1167–1171.

Benacerraf, P. (1973). Mathematical truth. *The Journal of Philosophy*, *70*(19), 661–679.

Brandom, R. B. (1994). *Making it explicit: reasoning, representing, and discursive commitment*. Cambridge: Harvard University Press.

Casullo, A. (2003). *A priori justification*. New York: Oxford University Press.

Field, H. (1980). *Science without numbers*. Princeton: Princeton University Press.

Frege, G. (1893-1903). *Grundgesetze der Arithmetik*, Volume I/II. Jena: Verlag Hermann Pohle

Goldberg, S. C. (Ed.). (2007). *Internalism and externalism in semantic and epistemology*. New York: Oxford University Press.

Jenkins, C. I. (2008). *Grounding concepts: An empirical basis for arithmetical knowledge*. New York: Oxford University Press.

Linnebo, Ø. (2012). Reference by abstraction. *Proceedings of the Aristotelian Society*, *112*(1), 45–71.

Maddy, P. (2011). *Defending the axioms on the philosophical foundations of set theory*. New York: Oxford University Press.

Peacocke, C. (2004). *The realm of reason*. New York: Oxford University Press.

Pincock, C. (2004). A new perspective on the problem of applying mathematics. *Philosophia Mathematica*, *3*(12), 135–161.

Pincock, C. (2010). Book review wandering significance: an essay on conceptual behavior. *Philosophia Mathematica*, *18*(3), 106–136.

Pincock, C. (2012). *Mathematics and scientific representation*. New York: Oxford University Press.

Putnam, H. (1983). *Realism and reason. philosophical papers* (Vol. 3). Cambridge: Cambridge University Press.

Quine, W. V. (1951). Two dogmas of empiricism. *Philosophical Review* LX(1): 20–43. In W. V. Quine, (Ed.), *From a logical point of view. Nine logico-philosophical essays*. (1953). 35–66. New York: Harper.

Shapiro, S. (2014). *Varieties of logic*. New York: Oxford University Press.

Wilson, M. (2006). *Wandering significance*. New York: Oxford University Press.

Part II
Realism in a World of Sets: from Classes to the Hyperuniverse

Chapter 7
Absolute Infinity in Class Theory and in Theology

Leon Horsten

How can I talk to you, I have no words

Virgin Prunes, *I am God*

Abstract In this article we investigate similarities between the role that ineffability of Absolute Infinity plays in class theory and in theology.

Keywords Absolute infinity · God · Reflection principles

7.1 Introduction

Zermelo held that there exist no collections beside sets. According to most interpretations—and I will go along with those here—he held that the mathematical universe forms a potentially infinite sequence of sets of a special kind, which he called ‘normal domains’. Quantification over sets is then necessarily restricted: we cannot quantify over all sets.

Versions of this article have been presented at various conferences and workshops, including at the *First Conference of the Italian Network for the Philosophy of Mathematics* (Milan 2014). I am grateful to the audiences at these events for invaluable questions, comments and suggestions. Among these, I am especially indebted to Anthony Anderson, Sam Roberts, Øystein Linnebo, Mark van Atten, Christian Tapp. Thanks also to two anonymous referees, for giving careful suggestions for improvement. But above all I am grateful to my colleague Philip Welch: I could never have written this article without the discussions that I have had with him about proper classes and reflection.

L. Horsten (✉)
University of Bristol, Bristol, UK
e-mail: Leon.Horsten@bristol.ac.uk

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_7

Cantor held that the set theoretic universe exists as a *completed* absolute infinity. The Burali-Forti paradox and Russell's paradox were initially interpreted as showing that Cantor's 'naive' set theory, as it is still sometimes called, is inconsistent. It was thought that Cantor had failed to recognise that the mathematical universe cannot itself constitute a set. Cantor himself protested that he never took the set theoretic universe as a whole to be a set. Nowadays, Cantor is rarely accused of having defended an outright inconsistent theory of sets. Nevertheless, according to the received view, Cantor's views about the set theoretic universe as whole are outdated, and ultimately philosophically untenable.[1]

Certainly there are, as we shall see, tensions in Cantor's view of the nature of the existence of the set theoretic universe. But a Cantorian viewpoint, when appropriately understood, is in some sense more powerful and fruitful than Zermelo's view of the set theoretic universe. This is manifested in the motivation of reflection principles in set theory. These are principles that state, roughly, that there are some sets in the universe that in certain ways resemble the universe as a whole (which is "too large" to be a set). The more the set theoretic universe resembles certain sets, the harder it becomes to distinguish it from them, i.e., the more ineffable it becomes. It is known that on Zermelo's conception of the set theoretic universe, at best only weak reflection principles can be motivated, which give rise to small large cardinal principles (Zermelo 1996). It is also known that a Cantorian conception of the set theoretic universe that countenances proper classes, can motivate somewhat stronger reflection principles (Bernays 1961; Tait 2005). We will see that a Cantorian viewpoint in fact motivates much stronger reflection principles, from which much stronger large cardinal axioms can be derived.

In western theology concepts of ineffability of the Absolutely Infinite have always played an important role. We will see that in some theological traditions, this ineffability is, like in class theory, understood in terms of indistinguishability of the Absolute (God) from certain entities that are in fact distinct from the Absolute. So there is a similarity between the role that the ineffability of Absolute Infinity plays in certain theological views and in class theory. The principal aim of the present article is to describe and explore this similarity.

7.2 Zermelo

Zermelo was the first to hold that, *Urelemente* aside, the mathematical universe consists only of sets. Through the work of Zermelo, Fraenkel and von Neumann, it became established in the 1920s that sets are governed by the laws of *ZFC*. This has become the most prevalent form of set theoretic platonism: there are only sets, and they obey the principles of *ZFC*.

---

[1] For one expression of this view, see Jané (1995).

The question then arises how the sets are related to the mathematical universe. Zermelo's viewpoint can arguably be canvassed as follows (Zermelo 1996, pp. 1231–1233). When we are engaged in set theory, our quantifiers always range over a domain of discourse D, which Zermelo calls a 'normal domain'. The entities over which our set theoretic quantifiers range are sets: they are governed by the principles of standard set theory (*ZFC*). Our domain of discourse D itself is also a collection. Since there are no collections other than sets (and *Urelemente*, for Zermelo, but we disregard them here), our domain of discourse must also be governed by the principles of *ZFC*. But, on pain of contradiction, D can then not be included as an element in our domain of discourse. Nonetheless, we can expand our domain of discourse so that it includes D as an element. The expanded domain of discourse D' can even be taken to be such that it also satisfies the principles of *ZFC*. But the expanded domain D' will again be a set. So the previous considerations apply to D' also: it cannot contain itself as an element, even though we can expand it further so as to remedy this defect. In sum, even though the domain of discourse can always be expanded, it never comprises all sets. The upshot is that for Zermelo, the mathematical universe is a potential infinite sequence of (actually infinite) domains of discourse that satisfy the principles of *ZFC*:

> What appears in one model as an 'ultrafinite non- or super-set' is in the succeeding model already a perfectly good, valid 'set' with a cardinal number and ordinal type, and it is itself a foundation-stone for the construction of a new domain. To the unlimited series of Cantor's ordinal numbers there corresponds a likewise unlimited double series[2] of essentially different set-theoretic models in each of which the whole classical theory is expressed. The two opposite tendencies of the thinking spirit, the idea of creative *advancement* and that of collective completion [*Abschluss*] […] are symbolically represented and reconciled in the transfinite number series based on well-ordering. This series in its unrestricted progress reaches no true completion; but it does possess relative stopping points, namely those 'limit numbers' which separate the higher from the lower models. (Zermelo 1996, p. 1233)

There are basic structural insights about the set theoretic universe that escaped Cantor. For instance, Zermelo in his later years viewed the set theoretic universe as structured into a layered hierarchy of initial segments V_α (with α ranging over the ordinals) that are sets. Zermelo even saw that there might be segments V_α that in a strong sense make all the axioms of *ZFC* true, as the quotation shows. Cantor did not see that far.

Zermelo's picture does raise some difficult questions. For one thing, it is not clear in which dimension the mathematical universe is supposed to vary. The notion of *creative advancement* suggests some form of progression or growth but it is not easy to see what the literal content of this metaphor really is. For another, there is the question how Zermelo can get his view across to us. (*What* is it that we cannot quantify over?) This does not imply that Zermelo's thesis about the essential restrictedness of quantification is false; however, it does seem difficult to see how this thesis can be communicated. I shall not pursue these worries here, but merely note that this can be taken as a reason for preferring Cantor's view over Zermelo's.

[2]Zermelo here means a series of initial segments V_α of the set theoretic universe V, for α ranging over the strongly inaccessible ordinals, and the membership relation restricted V_α.

7.3 Cantor on the Set Theoretic Universe

Cantor's theory of the nature of the set theoretic universe as a whole is not easy to summarise. His views seem to have undergone a transformation around 1895. I first discuss his earlier views, and then turn to his later views.

7.3.1 The Absolutely Infinite

Cantor's basic convictions preclude Zermelo's potential infinity of (completed) normal domains ever to be the final word about the nature of the set theoretic universe. The set theoretic universe could not, in Cantor's view, form a potential infinity of actual infinities because of what Hallett calls Cantor's *domain principle* (Hallett 1984, pp. 7–8), which says that every potentially infinite variable quantity presupposes an underlying fixed and completed domain over which the potentially infinite entity varies:

> There is no doubt that we cannot do without variable quantities in the sense of the potential infinite; and from this the necessity of the actual infinite can also be proven, as follows: In order for there to be a variable quantity in some mathematical inquiry, the 'domain' of its variability must strictly speaking be known beforehand through a definition. However, this domain cannot itself be something variable, since otherwise each fixed support for the inquiry would collapse. Thus, this 'domain' is a definite, actually infinite set of values. Thus, each potential infinite, if it is rigorously applicable mathematically, presupposes an actual infinite. (*Mitteilungen zur Lehre vom Transfiniten VII* (1887): (Cantor 1932, pp. 410–411), my translation)[3]

In particular, this means that even every absolute infinity of transfinite sets that potentially exists presupposes an actual, completed absolutely infinite domain as its range of variation.

Admittedly Cantor was in his writings not very explicit about what he did take the set theoretic universe as a whole to be. One problem is that it is not in every instance clear whether he has a theological or a mathematical conception of absolute infinity in mind. Indeed, he argues that it is the task not of mathematics but of 'speculative theology' to investigate what can be humanly known about the absolutely infinite (Cantor 1932, p. 378).[4] The following passage, for example, leans heavily to the theological side:

> I have never assumed a "Genus Supremum" of the acual infinite. Quite on the contrary I have proved that there can be no such "Genus Supremum" of the actual infinite. What

[3]In this quotation, Cantor speaks of the necessity of 'knowing' the domain of variation through a 'definition'. Surely Cantor is merely sloppy here, and we should discount the epistemological overtones. Another slip can be detected in Cantor's use of the word 'set' in this quotation: Cantor means the argument to be applicable not just to sets but also to absolute infinities. For a discussion of Cantor's sometimes sloppy uses of the term 'set', see Jané (2010), footnote 60.

[4]The connection between Cantor's conception of the mathematical absolutely infinite and his conception of God is explored in van der Veen and Horsten (2014).

> lies beyond all that is finite and transfinite is not a "Genus"; it is the unique, completely individual unity, in which everything is, which comprises everything, the 'Absolute', for human intelligence unfathomable, also that not subject to mathematics, unmeasurable, the "ens simplicissimum", the "Actus purissimus", which is by many called "God". (Letter to Grace Chisholm-Young (1908): (Cantor 1991, p. 454), my translation).

All this is related to the fact that in an Augustinian vein, Cantor takes all the sets to exist as ideas in the mind of God[5]:

> The transfinite is capable of manifold formations, specifications, and individuations. In particular, there are transfinite cardinal numbers and transfinite ordinal types which, just as much as the finite numbers and forms, possess a definite mathematical uniformity, discoverable by men. All these particular modes of the transfinite have existed from eternity as ideas in the Divine intellect. (Letter to Jeiler (1895): (Tapp 2005, p. 427), my translation)

Even though for this reason mathematical entities (sets and proper classes) are for Cantor not distinct from God, it is clear that he at times has a mathematical conception of the Absolutely Infinite in mind:

> The transfinite, with its wealth of arrangements and forms, points with necessity to an absolute, to the 'true infinite', whose magnitude is not subject to any increase or reduction, and for this reason it must be quantitatively conceived as an absolute maximum. (*Mitteilungen zur Lehre vom Transfiniten V* (1887): (Cantor 1932, p. 405), my translation)

This is the notion of absolutely infinite that I shall concentrate on in this article. I shall from now on disregard what Cantor takes to be the theological aspects of the mathematical absolutely infinite; I shall instead concentrate on Cantor's conception of the 'quantitatively absolute maximum', which is the set theoretic universe as a whole. From the passages discussed above, I conclude that he attributes to it the following properties. It is a fully determinate, fully actual ('completed'), inaugmentable totality. It is composed of objects (sets) that are of a mental nature ('ideas'). And unlike the sets in the mathematical universe, the universe as a whole cannot be uniquely characterised.

7.3.2 Inconsistent Multiplicities

From around the time when Burali-Forti published his 'paradox' (Burali-Forti 1897), one finds a subtle change of terminology in Cantor's writings. Whereas before, Cantor used the expression 'the Absolutely Infinite' for characterising the set theoretic universe, he now categorises the set theoretic universe and other proper classes (such as the class of all ordinals) as *inconsistent multiplicities*:

> If we assume the concept of a determinate multiplicity (of a system, of a realm ['Inbegriff'] of things), then it has proved to be necessary to distinguish two kinds of multiplicity (I always mean determinate multiplicities).

[5] For Cantor's most detailed account of the set theoretic universe in God's mind, see Tapp (2005), pp. 414–417. See also *Mitteilungen zur Lehre vom Transfiniten V*, footnote 3 (Cantor 1932, pp. 401–403).

> A multiplicity can be of such nature, that the assumption of the 'togetherness' ['Zusammenseins'] of its elements leads to a contradiction, so that it is impossible to conceive the multiplicity as a unity, as a 'finished thing'. I call such multiplicities absolutely infinite or inconsistent multiplicities. (Letter to Dedekind (1899), (Cantor 1932, p. 443), my translation)

Jané has argued that passages such as these indicate that Cantor no longer believed that the set theoretic universe forms a completed infinity (Jané 1995, Sects. 6 and 7). The strongest evidence for this thesis is perhaps the following quote from a letter from 1899 from Cantor to Hilbert:

> I am now used to call 'consistent' what before I referred to as 'completed', but I do not know if this terminology deserves to be maintained. (Letter to Hilbert (1899): (Cantor 1991, p. 399))

Jané speculated in Jané (1995) that instead of conceiving of the set theoretic universe as a completed whole, Cantor tacitly moved to a conception of the set theoretic universe as an irreducibly potential entity, whereby he arrived at a prefiguration of Zermelo's conception of the mathematical universe. This means that he must have by that time tacitly given up on the domain principle which says that every potential infinite has as its domain of variation an underlying completed infinite.

In his more recent (Jané 2010), Jané no longer claims that Cantor actually gave up the thesis of the existence of the mathematical universe as a completed infinity. But Jané rightly stresses that there remains a tension between Cantor's earlier commitments and Cantor's later terminology of inconsistent multiplicities:

> I submit that, owing to Cantor's allegiance to a changeless mathematical universe, Cantor's explanations [of the concept of inconsistent multiplicity] are indeed unconvincing. For how can the elements of a multiplicity fail to coexist if they all inhabit the same universe? (Jané 2010, p. 223)

And he thinks that the best way for Cantor to resolve this tension would be to embrace Zermelo's conception of the set theoretic universe as essentially open-ended.

Not everyone agrees with Jané's interpretation. It is true that Cantor's choice of words in the letter to Hilbert indicates that he no longer believed that the set theoretic universe can be mathematically understood as a whole. But the passages do not show that Cantor no longer believed that the set theoretic universe does not form an inaugmentable totality that forms the domain of our mathematical discourse past, present, and future. In Hauser's words:

> [B]y 'existing together' Cantor evidently means 'existing together as elements of a "finished" set'. Thus, what he is saying is merely that the totality of all transfinite numbers (or all alephs) does not constitute a set and therefore cannot be an element of some other set. But he is not denying that the transfinite numbers coexist in some other form, namely as *apeiron*, which is mathematically indeterminate, meaning that one cannot assign a cardinal or ordinal number to the totality of all numbers (Hauser 2013, Sect. 3).

The content of the notion 'apeiron' is notoriously unclear.[6] So this does not really help much in the clarification of the nature of the set theoretic universe. In other words,

[6]This notion goes back to Anaximander, and is variously translated as 'limitless','boundless', 'formless', 'the void'....

there is an unresolved interpretative debt at this point on the side of the defender of the Cantorian viewpoint. It seems that Jané is right that Cantor (or his defender) is facing a choice. Either she upholds Cantor's earlier view of the set theoretical universe and tries to make good philosophical sense of it, or she takes Cantor's characterisation of the mathematical universe as an inconsistent multiplicity as the final word, and tries to make sense of that. But both cannot be done at the same time.

What I propose to do is in the first instance to ignore Cantor's description of the set theoretic universe as an inconsistent multiplicity. In the following sections, I shall adopt Cantor's characterisation of the set theoretic universe as a completed whole, and discuss how it can be used to motivate what are called 'top down' reflection principles. Then I shall discuss a stronger reflection principle. We shall see that to make sense of this stronger reflection principle, elements both of Cantor's earlier views and elements of Cantor's later views on the nature of the set theoretic universe can be used.

7.4 Reflection

According to a time-honoured and influential view in the Judeo-Christian theological tradition, God is fundamentally ineffable. Cantor was well aware of this tradition and he extended it to mathematical absolutely infinites. After Cantor's time, in modern set theory, this view has been given *positive* expressions, which somewhat surprisingly have mathematical strength. These statements are known as *reflection principles*.

7.4.1 The Very Idea

The starting point of set theoretic reflection is the early Cantorian view that the mathematical absolutely infinite is unknowable:

> The Absolute can only be acknowledged, but never known, not even approximately known. (*Grundlagen einer allgemeinen Mannigfaltigkeitslehre* (1883), endnote to Sect. 4: (Cantor 1932, p. 205))

There are obvious connections with central themes in theology, especially with the medieval doctrine that only negative knowledge is possible of God (apophatic theology). As it stands, it is indeed a negative statement. However it can be given a positive interpretation as follows. Let us provisionally identify the mathematical absolutely infinite with the set theoretic universe as a whole (V). The universe V is unknowable in the sense that we cannot single it out or pin it down by means of any of our assertions: no true assertion about V can be made that excludes unintended interpretations that make the assertion true. In particular—and this is stronger than the previous sentence—no assertion that we make about V can ensure that we are talking about the mathematical universe rather than an object *in* this universe. So if we do make a true assertion ϕ about V, then there exist sets s such that ϕ is also true when it is interpreted in s.

In the late 1890s the Burali-Forti theorem made it abundantly clear that V is not the only actual whole that is absolutely infinite: the ordinals, for instance, form an absolutely infinite sequence. So in light of this we must say that the mathematical absolutely infinite comprises, in addition to the mathematical universe as a whole all other proper classes.[7] But in fact, the above argument should hold true for any proper class. Proper classes can then be said to be unknowable in the sense that no assertion in the language of sets can be true of *only* for some proper classes. So if we do make a true assertion concerning a proper class, then there exists sets about which this assertion is already true. If we truly describe mathematical absolute infinities, then there are set proxies for the absolute infinities such that our description can also truly be taken to range over the proxies.

Cantor did not explicitly articulate this line of argument. Yet he was probably the first one to make use of reflection as a principle motivating the existence of sets. He argues that the finite ordinals form a set because they can be captured by a definite condition:

> Whereas, hitherto, the infinity of the first number-class (I) alone has served as such a symbol [of the Absolute], for me, precisely because I regarded that infinity as a tangible or comprehensible idea, it appeared as an utterly vanishing nothing in comparison with the absolutely infinite sequence of numbers. (*Grundlagen einer allgemeinen Mannigfaltigkeitslehre* (1883), endnote to Sect. 4: (Cantor 1932, p. 205))

This can be seen as an application of a reflection principle.[8] Being closed under the successor operation is a set theoretic property of a mathematical absolute infinity (the ordinals). Reflection then allows us to infer that there must be a *set* that is closed under the successor operation, and hence that there must be a minimal such. This is the set ω. Anachronistically, one is tempted to say that Cantor is appealing to something like Montague-Levy reflection, which is a first-order reflection property that is provable in *ZFC*.[9]

7.4.2 Set Reflection

On the face of it, Zermelo's viewpoint uses a form of set theoretic reflection: every admissible domain of discourse in set theory is a 'normal domain', and this can by a reflective movement be seen to be a set. We cannot quantify over or in any way make use of proper classes, for, in his view, no such things exist. The set theoretic universe as a whole is not something we can talk about, according to Zermelo, for it

---

[7] Cantor's 1899 argument that the ordinals form an inconsistent totality is critically discussed in Jané (1995), pp. 395–396.

[8] Admittedly this passage is sufficiently vague as to be open to multiple interpretations. The view that this passage should be seen as an application of reflection is defended in Hallett (1984), pp. 117–118.

[9] The Montague-Levy reflection principle is discussed in Drake (1991), Chap. 3, Sect. 6.

never exists as a completed realm. So, literally speaking, Zermelo cannot, according to his own view, truly say that "*the* set theoretic universe is so rich that it contains many normal domains".

The best Zermelo can do is simply to postulate that above every ordinal, there is an ordinal which is the 'boundary number' of a normal domain. In modern terms, this is expressed as an axiom that postulates unboundedly many strongly inaccessible cardinals:

Axiom 1 $\forall \alpha \exists \beta : \beta > \alpha \wedge$ "β is a strongly inaccessible cardinal" .

This *seems* to say exactly what is required. It says that a fundamental property of the set theoretic universe, namely making the axioms of standard second-order set theory (ZFC^2) true,[10] is reflected in arbitrarily large set-sized domains. But closer inspection reveals that this cannot exactly be the case: there must be ordinal numbers that fall outside the quantifiers in this axiom. By Zermelo's own lights, the quantifiers in Axiom 1 must range over a domain of discourse that forms a set in a wider domain of discourse. There will be ordinals in this wider domain of discourse that do not belong to the 'earlier' domain of discourse.

Nonetheless, Axiom 1 and its relatives have some proof theoretic strength. They postulate the existence of 'small large cardinals' to which *ZFC* is not committed (Drake 1991, Chap. 4). That these large cardinals are still relatively small is witnessed by the fact that it is consistent for them to exist in Gödel's constructible universe L.

7.4.3 Class Reflection

Stronger reflection principles can be formulated if we take Cantor's idea of absolutely infinite multiplicities seriously. However, to study these reflection principles in a precise setting, logical laws governing them have to be formulated. The language that is assumed is the language of second-order (or two-sorted, if you will) set theory, where the membership symbol is expressing the only fundamental non-logical relation, and where we have two types of variables: the first-order variables range over sets ($x, y, \ldots$) and the second-order variables range over (proper and improper) classes ($X, Y, \ldots$). we shall from now on take the sets and classes to be governed by the principles of Von Neumann–Bernays–Gödel (NBG) class theory (and worry about the justification for this later).[11] Indeed, von Neumann's class theory, the precursor to Bernays' formulation of NBG, can be seen as a formalisation of Cantor's viewpoint (but not as a conceptual clarification).[12]

[10] Ranks V_α, for α strongly inaccessible, are models of ZFC^2.

[11] NBG differs from full ZFC^2 in that the second-order comprehension scheme is restricted to formulae that do not contain bound occurrences of second-order quantifiers.

[12] See von Neumann (1967).

If we take the point of view of Cantor's early theory of the mathematical universe, and take the point that there are more absolutely infinite collections than V alone, then we can express the reflection idea as follows:

Axiom 2 $\forall X : \Phi(X) \rightarrow \exists \alpha : \Phi^{V_\alpha}(X \cap V_\alpha)$,

where Φ^{V_α} is obtained by relativising all first- and second-order quantifiers to V_α and its power set, respectively, and where α does not occur free in Φ.

Zermelo's reflection principle (Axiom 1) only expresses that certain true class theoretical statements are reflected downwards (the axioms of ZFC^2). Axiom 2 states that *every* true (second order parametrised) class theoretic statement is reflected down to some set sized domain. Axiom 2 is stronger than Axiom 1: it implies large cardinal principles that postulate indescribable cardinals.[13]

Of course it is then natural to formulate reflection principles of orders higher than two in an analogous manner. However already the full third-order class reflection principle is inconsistent, at least for formulae that involve general parameters (Reinhardt 1974), (Koellner 2009, Sect. 3).[14] Third-order reflection restricted to a certain class of "positive" formulae is consistent and stronger than second-order class reflection (Tait 2005), but does not prove the existence of measurable cardinals or any other cardinals that are incompatible with $V = L$. Fourth-order reflection is inconsistent even when restricted to "positive" formulae (Koellner 2009, Sect. 5).

In sum, the situation is this. From Zermelo's conception of the set theoretic universe as a potential infinity of sets, the region of small large cardinals in the neighbourhood of inaccessible cardinals can be motivated. Due to its recognition of proper classes alongside of sets, the Cantorian point of view can be said to lead to the above stronger reflection principles of class reflection. However even those principles do not take us beyond the small large cardinal principles consistent with V being Gödel's constructible universe L. Indeed, the (tentative) conclusion of Koellner (2009) is that class theoretic reflection principles are either weak (in terms of large cardinal strength) or inconsistent.

7.5 Global Reflection

Gödel thought that *all* sound large cardinal principles can be reduced to reflection principles:

> 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

[13] Axiom 2 and its relatives were discussed in Bernays (1961). For a discussion of indescribable cardinals, see Drake (1991), Chap. 9.

[14] Parameter free sentences of higher orders are unproblematic.

> which either give some additional information or are not yet seen clearly to be derivable from the reflection principle as we understand it now (Wang 1996, 8.7.9).

This sentiment goes against the conclusions that Koellner reached, and is often regarded as implausible, because the familiar reflection principles are compatible with the principle that $V = L$. Nonetheless, we shall now argue that from a Cantorian point of view there may be more to Gödel's conjecture than is commonly thought.

Gödel himself was an adherent of Cantor's actualist viewpoint regarding the set theoretic universe rather than of Zermelo's potentialist viewpoint:

> To say that the universe of all sets is an unfinished totality does not mean objective undeterminateness, but merely a subjective inability to finish it. (Gödel, as reported in Wang (1996), 8.3.4)

We have seen that the set theoretic universe as a whole and all classes of sets are recognised by Cantor to (actually) exist: let us call this structure $\langle V, \in, \mathcal{C}\rangle$, where $\mathcal{C}$ contains all of Cantor's absolutely infinite collections. Then the reflection idea tells us that we cannot single this structure out by means of any of our assertions. Positively put, any assertions that hold in $\langle V, \in, \mathcal{C}\rangle$ must also hold in some set-size structure.[15]

There are various possible ways of trying to making this more precise. I shall not try to give a catalogue of the pro's and contra's of various options. Rather, I shall concentrate on one way that seems especially powerful, natural, and fruitful. As was mentioned earlier, it is to be assumed that $\langle V, \in, \mathcal{C}\rangle$ makes at least the principles of NBG true. Against this background, Welch proposed the following principle (Welch 2012):

Axiom 3 There is an ordinal κ and a nontrivial elementary embedding

$$j : \langle V_\kappa, \in, V_{\kappa+1}\rangle \longrightarrow \langle V, \in, \mathcal{C}\rangle$$

with critical point κ (i.e., $j(\kappa) > \kappa$ whereas below κ, j is the identity transformation).

This principle is called the *Global Reflection Principle* (*GRP*).[16] What the embedding function does is to act as the identity function on all elements of V_κ but to send the elements of $V_{\kappa+1}$ to elements of $\mathcal{C}$: $j(\kappa) = On$, $j(V_\kappa) = V$, $j(Card \cap \kappa) = Card, \ldots$ So Axiom 3 says that the set theoretic universe (with all its proper classes) is reflected in a particular way to a set-size initial segment of the universe.

The level of elementarity that is insisted upon can be varied. For the most part of this article we will require only Σ^0_1-elementarity where formulae are allowed to have class variables $X, Y, X, \ldots$ but are restricted to have only existential quantifiers which range over sets alone: we denote the resulting global reflection principle as $GRP_{\Sigma^0_1}$. But we could also insist on Σ^0_∞-elementarity or even Σ^1_∞-elementarity

[15] We will see later (Sect. 7.6.3) that the expression "any assertions" in this statement may need to be qualified.

[16] A philosophical defence of *GRP* is given in Horsten and Welch (forthcoming).

(denoted as $GRP_{\Sigma^0_\infty}$, $GRP_{\Sigma^1_\infty}$, respectively). Often, however, I will leave the level of elementarity required by the principle unspecified and simply speak of *GRP*.

The principle *GRP* says that the universe with its parts is, to a certain degree, indistinguishable from at least one of its initial parts V_κ and *its* parts. It says that the whole set theoretic universe with all its proper classes is mirrored in a set-sized initial segment $\langle V_\kappa, \in, V_{\kappa+1}\rangle$, where the first-order quantifiers range over V_κ, and where the reflection of a proper class C is obtained by 'cutting it off' at level V_κ.

GRP expresses the idea of reflection in a more powerful way than Axiom 2. Axiom 2 just says that each (second-order) statement is reflected from the set theoretic universe to some V_κ (where possibly different second-order statements are reflected in different V_κ's): therefore it does not entail that the universe as a totality particularly resembles any one *single* set-like initial segment. However *GRP* postulates that the whole universe $\langle V, \in, \mathcal{C}\rangle$ is indistinguishable from an initial 'cut' $\langle V_\kappa, \in, V_{\kappa+1}\rangle$ in a very specific way, namely in a way such that no large 'set' and no proper class can be distinguished from a proper subset of itself (its intersection with V_κ and with $V_{\kappa+1}$, respectively).

Even the weaker versions of *GRP* have strong large cardinal consequences. They entail the existence of sets that are incompatible with $V = L$, such as measurable cardinals, Woodin cardinals,…[17]

Thereby *GRP* is a more robustly ontological form of reflection than Axiom 2. In this respect, there is a striking connection with theological ideas that have a long history, as the following passage shows (Odo Reginaldus, quoted in Côté (2002), p. 78, my translation):

> How can the finite attain [knowledge of] the Infinite? On this question some said that God will show Himself to us in a mediated way, and that he will show Himself to us not in His essence, but in created beings. This view is receding from the aula…[18]

Of this passage, van Atten remarks (van Atten 2009, footnote 84, p. 22):

> From here it is only a small step to: "Suppose creature *A* has a perception of God. Then God is capable of making a creature *B* such that *A*'s perception cannot distinguish between God and *B*."

Indeed, I conjecture that the "view that is receding from the aula" to which Reginaldus is referring traces back to Philo of Alexandria, who writes in his *On Dreams*[19]:

> Thus in another place, when he had inquired whether He that is has a proper name, he came to know full well that He has no proper name, [the reference is to Exodus 6:3] and that whatever name anyone may use for Him he will use by licence of language; for it is not in the nature of Him that is to be spoken of, but simply to be. Testimony to this is afforded also by the divine response made to Moses' question whether He has a name, even "I am He that is (Exodus 3:14)." It is given in order that, since there are not in God things that man can

[17] The large cardinal strength of versions of *GRP* is discussed in Horsten and Welch (forthcoming) and in Welch (2012).

[18] "Quomodo potest finitum attingere ad infinitum? Propter hoc dixerunt alii quod deus contemperatum se exhibebit nobis, et quod ostendet se nobis non in sua essentia, sed in creatura".

[19] As quoted in Segal (1977), p. 163.

> comprehend, man may recognise His substance. To the souls indeed which are incorporeal and occupied in His worship it is likely that He should reveal himself as He is, conversing with them as friend with friends; but to souls which are still in the body, giving Himself the likeness of angels, not altering His own nature, for He is unchangeable, but conveying to those which receive the impression of His presence a semblance in a different form, such that they take the image to be not a copy, but that original form itself.

Although we have seen that Cantor was deeply familiar with the idea of God as ineffable, there is no textual evidence to suggest that he was familiar with theological literature in which the uncharacterisability of God is transformed into a *positive* principle, as was done in the passages above. Yet we have seen that Cantor at least once more or less explicitly made use of a mathematical reflection principle. But then it was done only in a fairly minimal way, namely, to argue for the existence of ω as a set. *GRP* is clearly a *much* stronger reflection principle than the one that Cantor implicitly appealed to in the quoted passage (Montague-Levy). But it is the class-theoretic counterpart of the theological thesis that is defended by Philo of Alexandria. Just as in the theological context, there are 'angels' such that every humanly describable property of God also applies to them, in the class theoretic context there are some sets such that every property of the universe also holds when relativised to them.

A key difference between the theological case and the class theoretic case is that we do not have a good theoretical understanding of the 'angels' in question, whereas we do have an excellent theory of sets. Also, in class theory, absolute infinities are reflected *in* the mathematical universe, whereas in the theological case, God is reflected in beings outside Himself.[20]

7.6 Sets, Parts, and Pluralities

Now that the philosophical motivation behind, and the content of, *GRP* has been explained, we turn to the ontological assumptions of the framework in which it is formulated.

7.6.1 GRP as a Second-Order Principle

So far we have only expressed *GRP* in a semi-formal way—in a manner of speaking often adopted by set theorists. If we formally want to express *GRP*, then at first blush it seems that we need a language of third order: the function j that is postulated to exist pairs sets of V_κ with themselves and sets of $V_{\kappa+1}$ with proper classes. Yet on the Cantorian perspective that we have adopted so far, only sets and collections of sets (proper classes) have been countenanced. But of course the mapping j that is

[20] I am indebted to an anonymous referee for these points.

postulated by *GRP* can in fact be *coded* as a second-order object: as a proper class K consisting of ordered pairs such that its first element a is in the domain of j (namely: $V_{\kappa+1}$) and the second element $j(a)$ is an element of $V_\kappa \bigcup \mathcal{C}$.

We also need a satisfaction predicate to express the elementarity of the embedding. *GRP* deploys two notions of truth: truth in the structure $\langle V_\kappa, V_{\kappa+1}, \in\rangle$, and truth in $\langle V, \mathcal{C}, \in\rangle$. Truth in $\langle V_\kappa, V_{\kappa+1}, \in\rangle$ can of course be defined in the language $\mathcal{L}^2_\in$. Σ^0_1-truth in $\langle V, \mathcal{C}, \in\rangle$ is also definable in $\mathcal{L}^2_\in$. But Σ^1_∞-truth is not (by Tarski's theorem). For the expression of the version of *GRP* that postulates Σ^1_∞-elementarity of j, it suffices to have a primitive Tarskian compositional satisfaction predicate T to $\mathcal{L}^2_\in$ and insist that the compositional truth axioms hold for $\mathcal{L}^2_\in$. This suffices to express what it means for a statement of $\mathcal{L}^2_\in$ to be true and to prove basic properties of truth. So, in sum, the fact that *GRP* postulates the elementarity of the embedding j even if we take a strong version of GRP that is Σ^1_∞ preserving, does not necessitate us to go up to third order.

7.6.2 Parts of V

As mentioned earlier, Cantor's distinction between sets and Absolute Infinities is a prefiguration of the distinction between sets and proper classes, which was articulated explicitly by von Neumann. The difference with Cantor's theory is that von Neumann did take classes as well as sets to be governed by mathematical laws. It is just that classes are objects *sui generis*: they obey different laws. Proper classes are objects that have elements, but they are not themselves elements. So, in particular, there is no analogue of the power set axiom for proper classes.

However it remained an open question how talk of proper classes ought to be interpreted. In particular, if proper classes are taken to be super-sets in some sense, then it is somewhat mysterious why they can have elements but not be elements. In Maddy's words (Maddy 1983, p. 122):

> The problem is that when proper classes are combinatorially determined just as sets are, it becomes very difficult to say why this layer of proper classes a top V is not just another stage of sets we forgot to include. It looks like just another rank; saying it is not seems arbitrary. The only difference we can point to is that the proper classes are banned from set membership, but so is the κth rank banned from membership in sets of rank less than κ.

And then why is there no singleton, for instance, that contains the class of the ordinals as its sole element? An alternative would be to say that proper classes can be collected into new wholes, but that these could (for obvious reasons) not themselves be proper classes. They would be again a *sui generis* kind of objects: super-classes. But in this way we embark on a hierarchical road that few find worth traveling. On this picture, classes, super-classes, et cetera, look too much like sets. We seem to be replicating the cumulative hierarchy of sets whilst incurring the cost of introducing a host of different kinds of set-like objects.

I propose instead to adopt a *mereological* interpretation of proper classes.[21] On this view, the mathematical universe is a mereological whole, and classes, proper as well as improper, are parts of the mathematical universe. We can identify those parts of V that are also parts of a set, i.e., that are set sized, with sets. The threat of a hierarchy of super- and hyper-wholes is not looming here. The fusion of the parts of a whole does not create a super-whole, but just the whole itself. So there is no mereological analogue of the creative force of the power set axiom.[22]

Such a mereological interpretation of *classes* is similar to David Lewis' interpretation of *sets* (Lewis 1991, 1993). Lewis takes sets to be generated by the singleton function and unrestricted mereological fusion. So sets have subsets as their mereological parts. Similarly, in the interpretation of classes that is proposed here, every class is a fusion of some singletons, and classes have sub-classes (and not their elements) as their proper parts. Also sets will have sub-sets as their parts; but in contrast to proper classes, they *are* also elements (of sets and of classes).

The difference between the proposed interpretation of the range of the second-order quantifiers and Lewis' theory of classes is that on the proposed interpretation set theory is taken as given. Lewis regards the relation between an entity and its singleton as thoroughly "mysterious" (Lewis 1991, Sect. 2.1). Reluctantly, he takes it to be a structural relation (Lewis 1991, Sect. 2.6). Derivatively, there is, in Lewis' approach, something mysterious about all sets. This is not the view that is taken here. I assume set theory from the outset, and do not commit myself to any specific interpretation (reductive or non-reductive) of the membership relation. Given the singleton-relation that is part of set theory, the elementhood relation for classes can be explained in a straightforward way. Explaining the element-relation for sets is outside the scope of this article.

The mereological interpretation of classes satisfies the two desiderata that according to Maddy an interpretation of class theory has to satisfy simultaneously (Maddy 1983, p. 123)[23]:

1. Classes should be real, well-defined entities;
2. Classes should be significantly different from sets.

The first desideratum is satisfied because classes are just as real and well-defined as sets. The second desideratum is satisfied because the laws of parthood are significantly different from the laws governing sets.

[21] See also Horsten and Welch (forthcoming).

[22] Even the Augustinian idea that sets are ideas in God's mind is compatible with this view. Within such a framework, the mereological conception of classes would result in conceiving of classes (proper and improper) as *parts* of God's mind.

[23] It seems to me that Maddy's own view of classes does not completely satisfy the first desideratum. The reason is that she takes the class membership relation to be governed by partial logic. According to her theory, there is in many cases no fact of the matter whether a given class is an element of another given class.

7.6.3 Mathematical and Mereological Reflection

Σ^0_∞ statements can be classified as as *mathematical* statements because they only quantify over sets. Σ^1_∞ statements can be classified as *mereological* (or in Cantorian vein one might say *theological*) statements because they quantify over proper classes, which on the proposed interpretation are regarded as extra- or supra-mathematical objects. In other words, we might call $GRP_{\Sigma^0_\infty}$ mathematical global reflection, whereas $GRP_{\Sigma^1_1}$ must already be regarded as mereological global reflection.

As mentioned before, already $GRP_{\Sigma^0_1}$ gives us strong large cardinal consequences. But one can go further than this and insist on mereological global reflection ($GRP_{\Sigma^1_1}$, for instance). In fact, even $GRP_{\Sigma^1_\infty}$ does not express the reflection idea in its strongest form. Recall that the guiding idea was that the set theoretic universe is *absolutely indistinguishable* from some set-like initial segment of V. *GRP* requires that the embedding j is elementary with respect to the second-order language of set theory *without the satisfaction predicate*. If we have a satisfaction predicate in our arsenal, we might require even stronger elementarity, viz. with respect to the second-order language *including the primitive satisfaction predicate*. Since this same satisfaction predicate is used to express the elementarity of j, it will have to be a non-Tarskian, type-free satisfaction predicate. For instance, one can contemplate using a truth predicate that is governed by axioms in the spirit of the Kripke–Feferman theory (Feferman 1991).

7.6.4 Names of God

In its most extreme form, negative theology states that no property can be truly predicated of God. In a more positive vein, one might say that everything that we can truly say about God, also holds for some being that is less exalted than God. (Note, once again, that this is not equivalent to the thesis in the first sentence of this paragraph.) Both these theories raise a difficult question, which was perhaps first articulated forcefully by Dionysius the Areopagite. If everything we truly say about God is also true about some angel(s), or if nothing we say about God is true at all, then how can we name God in the first place? What is it that makes our uses of the word 'God' refer to God rather than to angels in the first place?[24] The mirror image of this challenge for *GRP* is just as troublesome. If *everything* we truly say about V and $\mathcal{C}$ is also true about some set V_κ and its subsets, then what makes it the case that when we are using 'V' and '$\mathcal{C}$' in this article, these terms refer to the set-theoretic universe and its classes, respectively, rather than to some set and its subsets? If we insist on articulating *GRP* as requiring full Σ^1_∞ elementarity, then we only have the primitive notion of satisfaction to single out $\langle V, \mathcal{C}, \in\rangle$. However, if we articulate *GRP* as insisting only on a form of mathematical elementarity (Σ^0_1 elementarity or Σ^0_∞ elementarity), then this worry is not pressing. Then we can say that mereological

[24]See the quotation of Dionysius in Sect. 7.5.

(or “theological”) statements might allow us to distinguish V and $\mathcal{C}$ from every set together with its subsets.

There is a less philosophical reason for perhaps being hesitant to endorse mereological reflection. $GRP_{\Sigma^1_\infty}$ entails the axioms MK of Morse–Kelley class theory, i.e., ZFC^2 with its impredicative class comprehension axiom. MK holds at $(V_\kappa, V_{\kappa+1}, \in)$, and is then sent up by virtue of the Σ^1_∞ elementarity of j. So one might as well have started with the very “class-impredicative” theory MK as one's background theory.[25] But against this, one might say that on an actualist conception of V and its parts, for the same reasons that impredicative definitions of sets are unproblematic,[26] impredicative definitions of classes are unproblematic also.

7.7 In Closing

According to many, Cantor's early view of the mathematical universe as a whole is hopelessly entangled with his theological views (Tapp 2014). In contrast, his later view of the set theoretic universe and proper classes more generally as ‘inconsistent multiplicities’ is less so, and can be seen as a first step in the direction of a modern view of the set theoretic reality. It can then be seen as a prefiguration of a potentialist conception of the mathematical universe *à* la Zermelo (Jané 1995).

In this article I have argued that good non-theological sense can be made of Cantor's earlier view of the set theoretic universe. Sets are all the mathematical objects there are. All the sets together form, as the early Cantor said, a completed whole: the mathematical universe V. However V itself is not a *mathematical* object. Proper classes are parts of the universe. Every part of V is a completed whole. Every set is an element of V. The parthood relation corresponds to the subclass relation, which is a transitive relation. So parthood is not the same as membership, even for sets: not all sets are transitive. The language of sets and parts of V is the language of second-order set theory $\mathcal{L}^2_\epsilon$. The first-order quantifiers range over all sets. The second-order quantifiers range over all the parts of V. So we are ontologically committed to the existence of sets, the universe of all sets, and all of its parts: no further ontological commitments are made. The sets certainly satisfy *ZFC*. The parts of V satisfy at least predicative second-order comprehension. And the class replacement axiom also holds. So we are licensed to postulate NBG class theory in the language $\mathcal{L}^2_\epsilon$. If one takes a Gödelian stance towards impredicative definitions, then even impredicative second-order comprehension is acceptable. If that is so, then the axioms of Morse-Kelly class theory are motivated.

[25] This argument does not go through if instead j is only Σ^0_1 elementary: there is then not enough elementarity to preserve the impredicative second order comprehension scheme upwards. Nonetheless, since MK holds at $(V_\kappa, V_{\kappa+1}, \in)$, accepting $GRP_{\Sigma^0_1}$ still commits one to believing that impredicative second-order logic is at least coherent.

[26] See Gödel (1984).

Not only is this interpretation of Cantor's earlier view perfectly coherent. It is also mathematically fruitful. It allows us to indirectly motivate strong principles of infinity (large cardinal axioms). Large cardinal principles play an important role in contemporary set theory. However whereas the axioms of *ZFC* seem to be fairly generally accepted to hold of the set theoretic universe, there is no general agreement that most of the large cardinal principles hold.

Gödel argued that mathematical axioms can be motivated in two ways: intrinsically, and extrinsically (Gödel 1990). Extrinsic support for an axiom derives from its consequences. Thus extrinsic motivations are success arguments; they are instances of *Inference to the Best Explanation*. Many believe that intrinsic justification for mathematical principles is more reliable than extrinsic justification. Indeed, many do not think that external motivation for a mathematical axiom can provide strong confirmation of its truth (Tait 2001, p. 96). So it is an important question to what extent large cardinal principles can be motivated intrinsically.

Mathematical reflection principles are intrinsically motivated. These arguments follow a pattern of reasoning that has its roots in the Judeo-Christian theological tradition. This argument starts from the negative premise of the transcendence of God: there is no defining condition in any human language that is satisfied by Him and by Him alone. From this it follows that if we can truly ascribe a property to God, this property must hold of some entity that is different from God as well. This conditional positive statement can justly be called a *first theological reflection principle*. This argument can be strengthened if we assume the stronger negative premise that not even an infinite body of humanly describable conditions characterise God uniquely. This means that there must be an entity that is different from God and that satisfies all properties that can be truly ascribed to God. This then is a *second theological reflection principle*. We have seen how it was clearly articulated by Philo of Alexandria. The first theological reflection principle is the exact analogue of Bernays' second-order reflection principle. The second reflection principle is the analogue of the Global Reflection Principle.

Cantor did not see this far. On the theological side, there is no evidence that he was aware of statements of the second theological reflection principle. On the class theoretic side, he did not have the resources to even articulate the Global Reflection Principle: the cumulative rank structure of the set theoretic universe had yet to be discovered. We have seen that Cantor himself did (somewhat implicitly) appeal to a reflection principle on one occasion. But what he appealed to was a first-order reflection principle (Montague-Levy), and it is known that first-order reflection principles are provable in *ZFC*. In general, Cantor mostly referred to the epistemic transcendence of the set theoretic universe as a whole instead of focussing on its positive consequences (reflection principles).

The global reflection principle in its stronger forms is essentially a *second-order* reflection postulate. So to interpret it, we have to assign a clear meaning to the second-order quantifiers. On Zermelo's potentialist picture, this seems a tall order. Perhaps what Zermelo calls 'meta-set theory' allows quantification over absolute infinities, but Zermelo never clearly explained what he meant by 'meta-set theory'.

The pluralist interpretation of second-order quantification fares better (Uzquiano 2003). It may well give us a fairly clear interpretation of the second-order quantifiers. But on this interpretation, and therefore also on the interpretation of the second order quantifiers as ranging not only over sets but also over 'inconsistent multiplicities', the motivation for *GRP* becomes opaque. It is on this interpretation hard to make sense of the motivation for *GRP* in terms of a notion of resemblance. As far as I can see, it is only in terms of the interpretation of the second-order quantifiers as ranging over parts of the universe that the intrinsic motivation of *GRP* can be articulated. For this reason I conclude that the early Cantorian view of the set theoretical universe is mathematically the most fruitful one. Theology is not conservative over mathematics.

References

Bernays, P. (1961). Zur Frage der Unendlichkeitsschemata in der axiomatische Mengenlehre. In *Essays on the foundations of mathematics*, (pp. 3–49). Oceanside: Magnus Press.

Boolos, G. (1985). Nominalist platonism. *Philosophical Review*, *94*, 327–344.

Burali-Forti, C. (1897). Una questione sui numeri transfiniti. *Rendiconti del Circolo Matematico di Palermo*, *11*, 154–164.

Cantor, G. (1932). *Abhandlungen mathematischen und philosophischen Inhalts*. Herausgegeben von Ernst Zermelo: Verlag Julius Springer.

Cantor, G. (1991). *Briefe, Herausgegeben von Herbert Meschkowski und Winfried Nilson*. Berlin: Springer.

Côté, A. (2002). *L'infinité divine dans la théologie médiévale (1220–1255).* Berlin: Vrin.

Dionysius the Areopagite (1920). In C. E. Rolt (Ed.), *On the divine names and the mystical theology.* Christian Classics Ethereal Library.

Drake, F. (1991). *Set theory. An introduction to large cardinals.* North-Holland.

Feferman, S. (1991). Reflecting on Incompleteness. *Journal of Symbolic Logic*, *56*, 1–49.

Gödel, K. (1990). *What is Cantor's continuum problem?* In: Gödel, K., *Collected Works, Volume II: Publications 1938–1974* (pp. 254–270). Oxford: Oxford University Press.

Gödel, K. (1984). *Russell's mathematical logic.* Reprinted in: P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics: Selected readings* (2nd ed., pp. 447–469). Cambridge: Cambridge University Press.

Hallett, M. (1984). *Cantorian set theory and limitation of size.* Oxford: Clarendon Press.

Hauser, K. (2013). Cantor's Absolute in metaphysics and mathematics. *International Philosophical Quarterly*, *53*, 161–188.

Horsten, L., & Welch, P. (forthcoming). Absolute Infinity. *Journal of Philosophy*.

Jané, I. (1995). The role of the absolutely infinite in Cantor's conception of set. *Erkenntnis*, *42*, 375–402.

Jané, I. (2010). Idealist and realist elements in Cantor's approach to set theory. *Philosophia Mathematica*, *18*, 193–226.

Kanamori, A. (1994). *The higher infinite Large cardinals in set theory from their beginnings*. Berlin: Springer.

Koellner, P. (2009). On reflection principles. *Annals of Pure and Applied Logic*, *157*, 206–219.

Lewis, D. (1991). *Parts of classes*. Oxford: Basil Blackwell.

Lewis, D. (1993). Mathematics is megethology. *Philosophia Mathematica*, *3*, 3–23.

Maddy, P. (1983). Proper classes. *The Journal of Symbolic Logic*, *48*, 113–139.

Reinhardt, W. (1974). *Remarks on reflection principles, large cardinals, and elementary embeddings*. In: *Proceedings of Symposia in Pure Mathematics* (Vol. 10, pp. 189–205). Providence: American Mathematical Society.

Segal, A. (1977). *Two powers in heaven. Early rabbinic reports about christianity and gnosticism*. Brill.

Tait, W. (2001). Gödel's unpublished papers on the foundations of mathematics. *Philosophia Mathematica*, *9*, 87–126.

Tait, W. (Ed.). (2005). *Constructing cardinals from below.* In *The provenance of pure reason: Essays in the philosophy of mathematics and its history* (pp. 133–154). Oxford: Oxford University Press.

Tapp, C. (2005). *Kardinalität und Kardinäle: Wissenschaftshistorische Aufarbeitung der Korrespondenz zwischen Georg Cantor und katholischen Theologen seiner Zeit*. Franz Steiner.

Tapp, C. (2014). Absolute Infinity—a bridge between mathematics and theology? In H. Friedman & N. Tennant (Eds.), *Foundational adventures. Essays in honor of Harvey M. Friedman* (pp. 77–90). College Publications.

Uzquiano, G. (2003). Plural quantification and classes. *Philosophia Mathematica*, *11*, 67–81.

van Atten, M. (2009). Monads and sets. On Gödel, Leibniz, and the reflection principle. In G. Primiero & S. Rahman (eds.), *Judgement and Knowledge. Papers in honour of B.G. Sundholm.* College Publications, pp. 3–33.

van der Veen, J., & Horsten, L. (2014). Cantorian infinity and philosophical concepts of god. *European Journal for the Philosophy of Religion*, *5*, 117–138.

von Neumann, J. (1967). An axiomatization of set theory. In J. Van Heijenoort (Ed.), *From Frege to Gödel. A source book in mathematical logic (1879–1931).* Massachusetts: Harvard University Press.

Wang, H. (1996). *A logical journey. From gödel to philosophy*. Massachusetts: MIT Press.

Welch, P. (2012, June). *Global reflection principles.* Paper prepared for *Exploring the frontiers of incompleteness.* Isaac Newton Institute Pre-print Series, No. INI12050-SAS.

Zermelo, E. (1996). *On boundary numbers and domains of sets.* (M. Hallett, Trans.). In: W. Ewald (Ed.), *From Kant to Hilbert: A source book in mathematics* (Vol. 2, pp. 1208–1233). Oxford: Oxford University Press.

Chapter 8
Sets and Descent

Brice Halimi

Abstract Algebraic Set Theory, a reconsideration of Zermelo-Fraenkel set theory (ZFC) in category-theoretic terms, has been built up in the mid-nineties by André Joyal and Ieke Moerdijk. Since then, it has developed into a whole research program. This paper gets back to the original formulation by Joyal and Moerdijk, and more specifically to its first three axioms. It explains in detail that these axioms set up a framework directly linked to descent theory, a theory having to do with the shift from local data to a global item in modern algebraic geometry. Fibered categories, introduced by Grothendieck, provide a powerful framework for descent theory: They formalize in a very general way the consideration of local data of different kinds over the objects of some base category. The paper shows that fibered categories fit Joyal and Moerdijk's axiomatization in a natural way, since the latter actually aims to secure a descent condition. As a result, Algebraic Set Theory is shown to accomplish, not only an original and fruitful combination of set theory with category theory, but the genuine graft of a deeply geometric idea onto the usual setting of ZFC.

Keywords Algebraic set theory · Fibered categories · Descent theory

Several arguments have been put forward against category theory as a possible foundational theory for mathematics. But none is compelling. First, a central objection has been made to category theory, to the effect that the very notion of category (defined as a "collection" of objects together with a "collection" of arrows) presupposes the notion of set. Nothing prescribes, however, to conceive of an object or an arrow as being some element of a *set* in the sense of set theory. Otherwise, any formal language (based on some "collection" of symbols) would presuppose, if not formal set theory (ZFC, say), at least some prior theory of what a "collection" is. And this would apply to the language of that theory itself. In fact, category theory only

B. Halimi (✉)
Université Paris Ouest Nanterre La Défense (IREPH) & SPHERE, Paris, France
e-mail: bhalimi@u-paris10.fr

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_8

presupposes a background universe whose existence and set-theoretic constitution is implicit, but which has nothing to do with a model of ZFC. Set theory presupposes such a background universe as much as category does: Models of ZFC are lifted out of it as much as any category is.

Most mathematical notions are defined in set-theoretic terms. However, thinking in those terms is but confusing a mathematical theory with its set models, on the basis of Gödel's completeness theorem. The theory of fields, for instance, has nothing to do as such with set theory. So, even though the objects and arrows which compose a given category can be taken to be set-theoretic constructions, this is but one available option. Functorial semantics, and sketch theory in particular, is a nice illustration of that: A mathematical object such as a field can be introduced as a functor from a diagrammatic presentation of the field structure itself, to some realization category which is usually the category *Sets* of all sets and maps, but which can be other than *Sets*. Moreover, it is a more and more acknowledged fact that there are non-concrete categories, a category C being said to be *concrete* in case there exists a faithful functor from C to *Sets*. For instance, the homotopy category of the category of topological spaces, Ho(*Top*), is not concrete, as proved by Peter Freyd.[1] Higher categories, such as the category to which modern homotopy theory gives rise, provide other examples of non-concrete categories.

It is a fact that the category *Sets* is pervasive throughout mathematics, and functorial semantics cannot get rid of it, not in the least. Moreover, the Yoneda lemma itself proves that any category, however abstract it is, can be embedded into a category of presheaves of sets.[2] Nevertheless, *Sets* is brought up almost always merely as the "universal" or "free" category with colimits generated by one object (any singleton set). So an object of *Sets* has not to be construed as a set in the sense of classical set theory, and *Sets*, as a universal category, is characterized up to categorical equivalence only, which is a way to keep track of the generality of most mathematical constructions. Moreover, if you fix a base topos S and work with S-toposes (toposes over S), the set theory constituted by the internal language of S becomes the standard set theory. So ZFC is but one particular case of set theory. This led William Lawvere to extract "an elementary theory of the category of sets" in the form of category-theoretic "elementary" axioms, free of any reference to sets.[3] Tom Leinster, in a recent text, "Rethinking set theory," goes as far as to say: "The approach described here is not a rival to set theory: it *is* set theory".[4]

[1]See Freyd (2004).

[2]The Yoneda lemma applies only to a locally small category, and the notion of local smallness presupposes the notion of set, so sets should be given beforehand anyway. But, once the category of sets is assumed, then most mathematical constructions can be understood as pertaining to presheaves of sets, and thus as being pointwise set-theoretic in nature. And this is a result of category theory itself. I thank an anonymous referee for pointing this out to me.

[3]See MacLane and Moerdijk (1992) (p. 163) for a detailed exposition.

[4]Leinster (2012), p. 2.

This paper, however, does not seek to elicit arguments against ZFC, in favor of its replacement by an alternative category-theoretic framework. It rather claims that one should go beyond the contest, at least beyond too naively framed a contest between set theory and category theory, and that some mathematicians already went beyond it. Not that the contest should be left to mathematical practice to be settled. Indeed, the rivalry between set theory and category theory is undoubtably also a historical and sociological one, coming precisely from mathematical practice: Championing either set theory or category theory often implies that one extrapolates a tradition within mathematics (analysis and logical semantics in the case of set theory, algebraic geometry and modern algebraic topology in the case of category theory) to the whole of mathematics. Nevertheless, the rivalry between set theory and category theory does certainly not boil down to a mere historical or sociological issue. It has a deep mathematical core. This is precisely the reason why several mathematical theories feed upon the *combination* of set theory and category theory, which is much more fruitful than the endeavor to grant either theory the status of true foundations. This paper is devoted to a central, although too neglected, example of such a combination of both theories, namely Algebraic Set Theory (AST), a theory founded and built up almost from scratch in the nineties by two preeminent mathematicians, André Joyal and Ieke Moerdijk.

Set theory may have a peculiar, foundational status, but it constitutes also a specific branch of mathematics, and as such interacts with the other branches of mathematics. On that score, one can look at set theory from a non-set-theoretic point of view. This sounds very common (think of topos theory), but the purpose of AST, rather than replacing ZFC with something else, is to deepen ZFC by mixing it with a genuinely geometric perspective, where the shift from local data to a global item comes to the fore (this will become clearer in a while). Actually, AST can be described as a reconstruction of ZFC in a category-theoretic framework inspired by modern algebraic geometry—and more specifically, as will presently appear, by descent theory, which calls for the theory of fibered categories. This is what this paper will argue and explain. Focusing on the first three axioms of the original axiomatization of AST by Joyal and Moerdijk, it aims to show in detail that these axioms make sense essentially in reference to descent theory. One remarkable feature of AST is thus that it grafts the theory of fibered categories onto ZFC and thereby crossbreeds a new concept of set, or rather a new conception of what the very sets of ZFC, not the ones of some alternative topos, are. In doing so, AST goes beyond the unfruitful opposition often established between set theory and category theory. Furthermore, it challenges the linear order that would lead from foundational theories to specialized ones, and also holds out hopes of further applications of fibered categories to set theory and logic.

8.1 The First Axioms of Algebraic Set Theory

Up to now, AST has been explored along two main different approaches: a geometric one, embodied by the seminal work of Joyal and Moerdijk;[5] and a logical one, represented by Steve Awodey and Alex Simpson, among many others.[6] To put it briefly, whereas Joyal and Moerdijk seek to generalize models of ZF with the notion of "free ZF-algebra," in order to apply algebraic tools to models of set theory, the research program pursued by Awodey and his coworkers is to establish a detailed range of completeness results between different axiomatizations of set theory (intuitionistic set theory, predicative set theory, and so on), on the one hand, and, on the other, different collections of "categories of classes" (i.e., different collections of models of variants of AST). In what follows, I will be interested only in the former approach, and leave aside the latter.

8.1.1 Joyal–Moerdijk Axioms

Let C be a Heyting pretopos, which means that C is rich enough to interpret first-order logic and arithmetic.[7] (Still, C is not supposed to be a topos, and in particular to have "power objects.") The basic idea followed by Joyal and Moerdijk is to characterize a special subcollection of the collection of all arrows in C: a class S of "small maps" in C. Intuitively, an arrow in C is a small map if all its fibers have a set-like size. Then, an object X of C is said to be *small* if $X \to 1$ is a small map, which implies that X is set-like in size. Sets are now replaced with the small objects of C. But it is noteworthy that arrows are brought to the fore, instead of sets, and that sets themselves are primarily conceived of as "fibers."

The original motivation behind AST is the remark that, in the internal language of any topos, unbounded quantification is by construction excluded: Only quantification restricted to an object is available. In that perspective, it makes sense to formalize, within any topos, the distinction between set-like objects and class-like objects, so that class-wide quantification can be emulated.

The first six axioms for a class S of small maps of C are as follows:[8]

1. Any isomorphism of C belongs to S and S is closed under composition.
   This axiom corresponds to the fact that the union of a small family of small objects is a small object.

[5] Joyal and Moerdijk (1995).

[6] For an overview, see Awodey et al. (2014).

[7] See the Appendix B of Joyal and Moerdijk (1995) for details about Heyting pretoposes. See MacLane (1997) and Johnstone (2002) for all general notions of category theory mentioned in this paper.

[8] Joyal and Moerdijk (1995), pp. 7–8.

2. Stability under change base: For any pullback

$$\begin{tikzcd} Y' \arrow[r] \arrow[d, "g"'] \arrow[dr, phantom, very near start, "\lrcorner"] & Y \arrow[d, "f"] \\ X' \arrow[r, "p"'] & X \end{tikzcd},$$

if f belongs to S, so does g.
3. "Descent:" For any pullback along an epimorphic arrow p

$$\begin{tikzcd} Y' \arrow[r] \arrow[d, "g"'] \arrow[dr, phantom, very near start, "\lrcorner"] & Y \arrow[d, "f"] \\ X' \arrow[r, two heads, "p"'] & X \end{tikzcd},$$

if g belongs to S, so does f.
4. The arrows $0 \to 1$ and $1 + 1 \to 1$ belong to S.
This axiom, together with the axioms 1 and 5, ensures that the initial object is small and that any finite object is small.
5. If two arrows $f : Y \to X$ and $f' : Y' \to X'$ belong to S, then so does their sum $f + f' : Y + Y' \to X + X'$.
This axiom ensures that the disjoint union of two small objects is a small object.
6. In any commutative diagram

$$\begin{tikzcd} Z \arrow[rr, two heads, "p"] \arrow[dr, "g"'] & & Y \arrow[dl, "f"] \\ & B & \end{tikzcd}$$

with g epimorphic, if g belongs to S, so does f.

The above axioms are not all the axioms of AST. Three other axioms[9] must be added to define a class of small maps in C. But we will not discuss the latter, and will focus on the first three axioms only.

8.1.2 ZF-Algebras

In what follows, I will briefly indicate how the above-mentioned axioms catch up with ZFC.

[9] Axioms "**(A7)**," "**(S1)**" and "**(S2)**" (Joyal and Moerdijk 1995, pp. 8–9).

If L is a partially ordered set ("poset") in C, then $\mathrm{Hom}_C(A, L)$ is itself a poset in C. Now, given a particular poset L in C, for any maps $g : B \to A$ and $\lambda : B \to L$ in C, the *supremum of* λ *along* g is the map $\mu : A \to L$ (defined up to isomorphism) such that, for any $t : L' \to A$ and any $\nu : L' \to L$,

$$\begin{tikzcd}
L' \times_A B \arrow[r, dashed, "p_2"] \arrow[d, dashed, "p_1"'] & B \arrow[d, "g"] \arrow[dr, "\lambda"] & \\
L' \arrow[r, "t"] \arrow[rr, bend right, "\nu"'] & A \arrow[r, dashed, "\mu"] & L
\end{tikzcd}$$

$$\lambda \circ p_2 \leq \nu \circ p_1 \text{ in } \mathrm{Hom}_C(L' \times_A B, L) \text{ iff } \mu \circ t \leq \nu \text{ in } \mathrm{Hom}_C(L', L).$$

The poset L is said to be *S-complete* if any map to L has a supremum along any map in S.

Definition 1 A ZF-*algebra* in C is an S-complete poset L with all joins, endowed with a map $s : L \to L$.

The partial order $\leq$ of L corresponds to inclusion, and the map s corresponds to the singleton operation. Given a ZF-algebra $\langle L, s\rangle$ in C, it is possible to define a "membership relation" $\epsilon \rightarrowtail L \times L$ on L, by setting: $x \,\epsilon\, y$ iff $s(x) \leq y$ for any x and y "in" L. A notion of homomorphism of ZF-algebras in C is also readily available. As a consequence,[10] any object A of C generates a *free* ZF-*algebra* in C, written $V_{\langle C,S\rangle}(A)$, and called "the cumulative hierarchy on A".

Fact (Joyal and Moerdijk (1995), Theorem 5.5) *Adding two further assumptions about S, one has that* $A + V_{\langle C,S\rangle}(A)$ *is a model of* ZF *set theory with atoms.*

Models of ZF, including the cumulative hierarchy based on a set or a forcing extension, can thus receive a uniform treatment that makes them appear all as free algebras generated by a given basis.

8.2 Descent Theory

8.2.1 The Concept of Fibered Category

The main purpose of this paper is to explain the elliptic phrase "Descent" used to describe the third axiom of AST. Descent theory is a central tool coming from algebraic geometry, which unifies number-theoretic methods linked to Galois theory and topological methods concerning coverings and the construction of a space by

[10] Joyal and Moerdijk (1995), p. 30.

"glueing" pieces of it (see below). Since Grothendieck's work, descent theory has been developed in the very general context of "fibered categories," a.k.a. "fibrations," and this context precisely underpins the link between descent theory and AST. So I will explain, in order: what a fibered category is; what descent consists in, w.r.t. some given fibered category; which is the rationale of the first three axioms of AST, in the light of descent theory.

Definition 2 (Jacobs (1999), 1.1) A *fibered category* is the category-theoretic generalization of the notion of surjective map. It is a functor $p : F \to C$ such that, for any object $x \in F$ and any arrow $u : X' \to p(x)$ in C, there exists an arrow $\widehat{u} : x' \to x$ in F, "the Cartesian lift of u," lying above u (i.e., $p(\widehat{u}) = u$) and universal among all the arrows in F above u. In other words, the Cartesian lift $\widehat{u}$ of u is defined (up to isomorphism) by the two following conditions: (i) $p(\widehat{u}) = u$; (ii) for any arrow $g : y \to x$ in F, if there is $w : p(y) \to X'$ such that $p(g) = u \circ w$, then there is a unique arrow $h : y \to x'$ in F above w (i.e., $p(h) = w$) such that $g = \widehat{u} \circ h$, as in the diagram below:

$$\begin{tikzcd} & y \arrow[dl, dotted, "h"'] \arrow[dr, "g"] & \\ x' \arrow[rr, "\widehat{u}"] & & x \end{tikzcd}$$

$$\begin{tikzcd} & p(y) \arrow[dl, "w"'] \arrow[dr, "p(g)"] & \\ X' \arrow[rr, "u"] & & p(x) \end{tikzcd} .$$

The category C is called the "base category," F the "total space," and, for each object X of C, the category $F_X := p^{-1}(X)$ is a subcategory of F, called the "fiber" above X, which generalizes the set-theoretic notion of pre-image. The objects of F_X are all the objects x of F such that $p(x) = X$, and the arrows in F_X are all the arrows $f : x \to x'$ in F such that $p(f) = 1_X$ (the identity morphism on X).

The fundamental extra-ingredient carried by a fibered category, in comparison with surjective maps and pre-images, is the fact that the base category is endowed with some structure (as embodied by its arrows), and that the existence of a fibered category, owing to the correspondence $u \mapsto \widehat{u}$, requires a systematic connection between the relations between any two objects in the base category (arrows u), and the relations between the corresponding fibers in the total space (Cartesian lifts $\widehat{u}$).

A simple example of a fibered category is provided by set-indexed families of objects. The base category is the category *Sets* of all sets, and the total space is the category whose objects are all I-indexed families $(A_i)_{i\in I}$ and whose arrows $(A_i)_{i\in I} \to (B_j)_{j\in J}$ are all pairs $(u, (f_i)_{i\in I})$, where $u : I \to J$ is a map between the index sets, and each f_i is a map $A_i \to B_{u(i)}$. The fiber above any set I is the category Fam_I of all families indexed by I. This means that one considers the fibered category π that maps any set-indexed family $(A_i)_{i\in I}$ to I, and any arrow $(u, (f_i)_{i\in I})$ to u. The functor π is obviously surjective on objects, but it is surjective on arrows as well: Indeed, for a given family $(A_j)_{j\in J}$, any arrow $u : I \to J$ induces an inclusion arrow $(A_{u(i)})_{i\in I} \hookrightarrow (A_j)_{j\in J}$, where $(A_{u(i)})_{i\in I}$ appears to be the result of reindexing $(A_j)_{j\in J}$.

Given a fibered category $p : F \to C$, the choice of a certain Cartesian lift $\widehat{u} : x' \to x$ in F, for each pair $(x, u : X' \to p(x))$ of an object x of F and an arrow u in C, is called a *cleavage* of the fibered category p. (From now on, all fibered categories will be assumed to be equipped with a cleavage.) Writing $u^*(x)$ instead of x', that is, $\widehat{u} : u^*(x) \to x$, one can check that each arrow $u : X' \to X$ in C gives rise to a "base-change" functor $u^* : F_X \to F_{X'}$ between the corresponding fibers (in the reverse direction). The functor u^* is defined on arrows as follows: For any arrow $f : x \to x_1$ in F_X, $p(f \circ \widehat{u}_x) = 1_{X'} \circ u$ so, by definition of $\widehat{u}_{x_1}$,[11] there exists a unique arrow $u^*(f) : u^*(x) \to u^*(x_1)$ such that $p(u^*(f)) = 1_{X'}$ and $\widehat{u}_{x_1} \circ u^*(f) = f \circ \widehat{u}_x$, so that the diagram

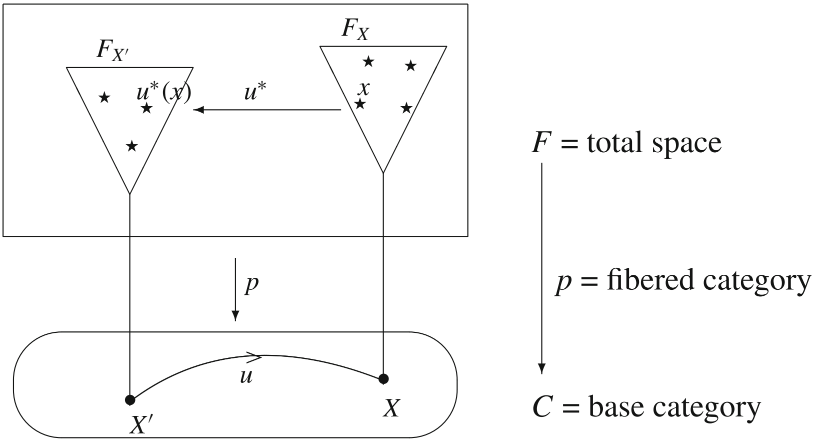
$$\begin{tikzcd} u^*(x) \arrow[r, "\widehat{u}_x"] \arrow[d, dashed, "u^*(f)"'] \arrow[dr, "f\circ\widehat{u}_x"] & x \arrow[d, "f"] \\ u^*(x_1) \arrow[r, "\widehat{u}_{x_1}"'] & x_1 \end{tikzcd}$$

commutes. The existence of base-change functors between fibers is another way to express the fact that the structure of the base category C is reflected in a principled way by the organization of the fibers in the total space F.

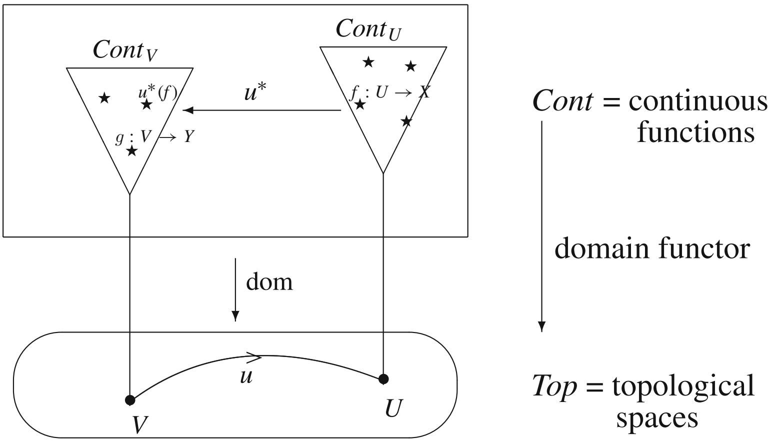
The mapping which assigns to each object X of C its fiber F_X, and to each arrow $u : X' \to X$ the corresponding functor $u^* : F_X \to F_{X'}$, constitutes a "pseudo-functor" from the category C to the category *Cat* of all small categories, called an *indexed category*. An indexed category can be thought of as a certain presentation of a fibered category, so that the definition of base-change functors amounts to the recognition of a fibered category.[12] Here is a picture of a fibered category in general:

[11] See condition (ii) of Definition 2 above.

[12] See Jacobs (1999), 1.4.4 (pp. 50–51) about the notions of pseudo-functor and of indexed category. The recovering of a fibered category (with a cleavage) from an indexed category is called the "Grothendieck construction;" See Jacobs (1999), 1.10 (pp. 107–108) for details.



Continuous functions provide an example of fibered category (over the category of topological spaces):



$$\text{For } f : U \to X, \quad \mathbf{u}^*(\mathbf{f}) = \mathbf{f} \circ \mathbf{u} : V \to X$$

The continuous functions between topological spaces are the arrows in the category of topological spaces, but the objects of a category *Cont* as well. The functor that sends each continuous function to its domain and acts on the arrows of *Cont* (to be defined, see below) in a corresponding way is an example of fibered category. This example can be generalized. Indeed, given a category C, let $C^{\rightarrow}$ be the category whose objects are all the arrows in C. Given two objects of $C^{\rightarrow}$, namely two arrows $f_1 : Y_1 \to X_1$ and $f_2 : Y_2 \to X_2$ in C, an arrow in $C^{\to}$ from f_1 to f_2 is a pair $(g' : Y_1 \to Y_2, g : X_1 \to X_2)$ of arrows in C such that the following diagram

$$\begin{tikzcd} Y_1 \arrow[r, "g'"] \arrow[d, "f_1"'] & Y_2 \arrow[d, "f_2"] \\ X_1 \arrow[r, "g"'] & X_2 \end{tikzcd}$$

commutes. The *domain functor* is the functor dom which assigns to each object $f : Y \to X$ of $C^{\to}$ (i.e., to each arrow in C) its domain Y, and to each arrow $(g', g) : f_1 \to f_2$ in $C^{\to}$, the arrow g' in C. It is a fibered category. Indeed, the fiber above each object X of C is the full subcategory $\mathrm{dom}^{-1}(X)$ of $C^{\to}$ composed of all the arrows $f : X \to Y$ in C with domain X, and the base-change functor u^*, for any arrow u in C, simply is pre-composition with u (i.e., $u^*(f) = f \circ u$). The example just above corresponds to the particular case where C is *Top*.

The symmetrical functor will bring us closer to AST. The *codomain functor*, cod : $C^{\to} \to C$, assigns to each object $f : Y \to X$ of $C^{\to}$ (i.e., to each arrow in C) its codomain X, and to each arrow $(g', g) : f_1 \to f_2$ in $C^{\to}$, the arrow g in C.

Fact (Vistoli (2005) (3.15), Jacobs (1999) (1.1.6)) *For any category C with pullbacks, the codomain functor* cod : $C^{\to} \to C$ *is a fibered category.*

The fiber $(C^{\to})_X$ above each object X of C simply is the full subcategory of $C^{\to}$ composed of all the arrows $f : Y \to X$ in C with codomain X; It is written C/X. Specifically, and importantly, the base-change functor u^* is given by pullback: For each arrow $u : X' \to X$ in C, u^* maps any $f : Y \to X$ in C/X to the pullback $u^*(f)$ of f along u. In other words, the diagram

$$\begin{tikzcd} X' \times_X Y \arrow[r, dashed] \arrow[d, dashed, "u^*(f)"'] \arrow[dr, phantom, "\lrcorner", very near start] & Y \arrow[d, "f"] \\ X' \arrow[r, "u"'] & X \end{tikzcd}$$

induces the mapping

$$\begin{tikzcd} u^*(f) \in C/X' & f \in C/X \arrow[l, mapsto, dashed, "u^*"'] \\ X' \arrow[r, "u"'] & X \end{tikzcd} .$$

Any collection S of arrows in a category C can be viewed as a subcategory of $C^{\to}$, namely the full subcategory of $C^{\to}$ whose collection of objects (arrows in C) is exactly S. (The collection S and the full subcategory of $C^{\to}$ associated to it will henceforth be identified.) The main point of the first two axioms of AST is the following.

Definition 3 Given a fibered category $p : F \to C$, a *fibered subcategory of* p is a fibered category $q : G \to C$ such that G is a subcategory $i : G \hookrightarrow F$ of F, $q = p \circ i$ (i.e., q is the restriction of p to G) and all the Cartesian lifts w.r.t. q are Cartesian lifts w.r.t. p.

Lemma 1 *Let C be a Heyting pretopos. If a full subcategory S of $C^{\to}$ verifies the Axioms 1 and 2 of a class of small maps (in the sense of* AST*), then the restriction of the codomain functor to S, $\mathrm{cod}_S : S \to C$, defines a fibered subcategory of $\mathrm{cod} : C^{\to} \to C$.*

Proof Axiom 1 says that all isomorphisms in C belong to S, which implies that, for any object X of C, 1_X belongs to S, and thus that there is at least one object (arrow) in S above each object of C. So the category S can really be said to be "above" C. Axiom 2 then says that any pullback of any arrow in S belongs itself to S. As Cartesian lifts w.r.t. the codomain functor are given by pullbacks, this exactly ensures that cod_S is a fibered subcategory of cod. □

8.2.2 Glueing Conditions

Descent theory constitutes an abstract framework geared to describing glueing processes, i.e., the shift from local data to a global item. Here is a typical example.[13] Let $(U_i)_{i\in I}$ be a covering of a topological space U (that is, a family of open subsets of U such that $\bigcup_{i\in I} U_i = U$), and suppose that for each $i \in I$ a continuous function $f_i : U_i \to V$ is given, in such a way that

$$\forall i, j \in I \;\; f_i|U_{ij} = f_j|U_{ij},$$

where $U_{ij} := U_i \cap U_j$. Then, obviously, there exists a unique function $f : U \to V$ such that $\forall i \in I \;\; f|U_i = f_i$.

This result can be framed and reformulated in the context of the domain functor $\mathrm{dom} : \mathit{Cont} \to \mathit{Top}$. Indeed, let's write $U' := \coprod_{i\in I} U_i$. The canonical map $u : U' \to U$ is both continuous and surjective. Moreover, the family $(f_i)_{i\in I}$ exactly constitutes an object in $\mathit{Cont}_{U'}$. Now, let's introduce the following pullback:

$$\begin{tikzcd} U' \times_U U' \arrow[r, dashed, "p_2"] \arrow[d, dashed, "p_1"'] \arrow[dr, phantom, "\lrcorner", very near start] & U' \arrow[d, "u"] \\ U' \arrow[r, "u"'] & U \end{tikzcd} .$$

[13] See Vistoli (2005), pp. 67–68.

In this diagram:

- $U' \times_U U'$ is a subset of $U' \times U'$, p_1 is the canonical projection onto the first factor, and p_2 the canonical projection onto the second factor.
- $U' \times_U U'$ corresponds to $\coprod_{i,j \in I} U_{ij}$. Indeed, it is the set of all $(x, y) \in U' \times U'$ such that $(u \circ p_1)(x) = (u \circ p_2)(y)$; equivalently, the set of all (x, y) with $x \in \coprod_{i \in I} U_i$ (say $x \in U_{i_1}$), $y \in \coprod_{i \in I} U_i$ (say $y \in U_{i_2}$) and $u(x) = u(y)$, this latter condition meaning that $x = y \in U_{i_1} \cap U_{i_2}$.
- For $x \in U' \times_U U'$, say $x \in U_{ij}$ for some i and some $j \in I$, $p_1(x)$ is x as an element of U_i, whereas $p_2(x)$ is x as an element of U_j.
- For $(f_i)_{i \in I} \in Cont_{U'}$, $p_1^*((f_i)_i) = (f_i|U_{ij})_{i,j}$ and $p_2^*((f_i)_i) = (f_i|U_{ji})_{i,j} = (f_j|U_{ij})_{i,j}$.

So, $(U_i)_i$ being represented by U', and $(U_i \cap U_j)_{i,j}$ by $U' \times_U U'$, the hypothesis $\forall i, j \in I \;\; \phi_{ij} : f_i|U_{ij} = f_j|U_{ij}$ can be rephrased as follows:

$$p_1^*((f_i)_i) = p_2^*((f_i)_i) \;\text{ in } Cont_{U' \times_U U'}.$$

More generally, replacing identity with an isomorphism, one assumes an isomorphism

$$\phi : p_1^*((f_i)_i) \simeq p_2^*((f_i)_i) \;\text{ in } Cont_{U' \times U'},$$

i.e., a family

$$\phi_{ij} : p_1^*(f_i) \simeq p_2^*(f_i)$$

of "glueing isomorphisms" in the fibers above the U_{ij}'s—those isomorphisms being compatible when they overlap, that is: $\phi_{ik} = \phi_{jk} \circ \phi_{ij}$. As soon as such isomorphisms are given, it becomes possible to glue the local data above the components U_i of the covering U', so as to get a global item above $U = \bigcup_i U_i$, namely a function $f : U \to V$ such that $\forall i \in I \;\; f|U_i \simeq f_i$.

This typical example leads to the following fully general formulation, entirely due to Grothendieck,[14] of glueing conditions over a category C, in the context of a fibered category with base C. To gain a sound understanding of what follows, it is important to recall that the cleavage of a fibered category $p : F \to C$ amounts to the choice, for each arrow $f : X'' \to X'$ in C (for instance a projection p_1 or p_2), of a certain Cartesian lift $\widehat{f} : f^*(x) \to x$ in F, and that this choice gives rise to a base-change functor $f^* : F_{X'} \to F_{X''}$ for each such f.

Definition 4 Let $p : F \to C$ be a fibered category (with a cleavage). Then, given an arrow $u : X' \to X$ in the base category C and $\xi' \in F_{X'}$, a *descent datum on* $\xi' \in F_{X'}$ is an isomorphism

$$\phi : p_1^*(\xi') \simeq p_2^*(\xi') \;\text{ in } F_{X' \times_X X'}$$

[14] Cf. Grothendieck (1962).

satisfying the following compatibility condition: $p_{31}^*(\phi) = p_{32}^*(\phi) \circ p_{21}^*(\phi)$, where each $p_{ij} : X' \times_X X' \times_X X' \to X' \times_X X'$ is the canonical projection onto the ith and jth components.

Here is the corresponding picture:

$$\begin{tikzcd}
p_1^*(\xi') \arrow[d, "\phi"', "\wr"] \arrow[dr] & & & \\
p_2^*(\xi') \arrow[r] & \xi' \arrow[r, squiggly] & ? & \text{items in } F \\
X' \times_X X' \arrow[r, shift left, "p_1"] \arrow[r, shift right, "p_2"'] & X' \arrow[r, "u"] & X & \text{items in } C \\
{[(U_{ij})_{i,j}]} & {[(U_i)_i]} & {[U]} & \text{[typical example]}
\end{tikzcd}$$

Intuitively, ξ' represents a collection of objects, each one above one of the components of a covering of X, and the isomorphism ϕ witnesses the agreement of all these objects over all the overlappings of the corresponding components of the covering. The question is, then: Can we think of ξ' as the collection of all the restrictions (to all the different components of the covering of X) of a *single* object above X? Since a descent datum ϕ can be viewed as a virtual glueing, the question is: *Is any virtual glueing over a covering of X actualizable as an object above X*? Given any arrow $u : X' \to X$, every object ξ above X induces an object $u^*(\xi)$ above X' which is naturally equipped with a descent datum, since $u^*(\xi)$ verifies: $p_1^*(u^*(\xi)) \simeq (u \circ p_1)^*(\xi) = (u \circ p_2)^*(\xi) \simeq p_2^*(u^*(\xi))$. The central question of descent theory pertains to the converse: Does any object above X', equipped with a descent datum, amount to an object above X? When such a converse holds, the arrow u is said to be a "strict descent morphism" (see below).

Definition 5 Given a fibered category $p : F \to C$, any arrow $u : X' \to X$ in C gives rise to the following pullback diagram, already mentioned:

$$\begin{tikzcd}
X' \times_X X' \arrow[r, dashed, "p_2"] \arrow[d, dashed, "p_1"'] \arrow[dr, phantom, "\lrcorner", very near start] & X' \arrow[d, "u"] \\
X' \arrow[r, "u"'] & X
\end{tikzcd} .$$

The arrow u is a *p-descent morphism* iff, for any two objects ξ, η of F_X, the diagram

$$\mathrm{Hom}_{F_X}(\xi, \eta) \xrightarrow{u^*} \mathrm{Hom}_{F_{X'}}(u^*(\xi), u^*(\eta)) \overset{p_1^*}{\underset{p_2^*}{\rightrightarrows}} \mathrm{Hom}_{F_{X' \times_X X'}}((u \circ p_1)^*(\xi), (u \circ p_2)^*(\eta))$$

is exact, in other words iff:

(A) For any two objects ξ, η of F_X, $u^* : F_X \to F_{X'}$ induces a 1-1 map from $\mathrm{Hom}_{F_X}(\xi, \eta)$ to the subset of $\mathrm{Hom}_{F_{X'}}(u^*(\xi), u^*(\eta))$ consisting of all the arrows $g : u^*(\xi) \to u^*(\eta)$ such that $p_1^*(g) = p_2^*(g)$.

The arrow u is a *strict p-descent morphism* (a.k.a. an *effective p-descent morphism*) iff, moreover:

(B) Giving an object of F_X exactly amounts to giving an object of $F_{X'}$ equipped with a descent datum: Any object of $F_{X'}$ equipped with a descent datum is isomorphic to some $u^*(\xi)$, for some $\xi \in F_X$.

Otherwise put, u is a strict p-descent morphism if whatever virtually glues actually glues along u. Indeed, $p_1^*(g) = p_2^*(g)$ means that all the components of g have identical restrictions to $X' \times_X X'$. The condition (A) says that, in such a case, the components of g can actually be glued together along u, so as to come from one single arrow in F_X: $g = u^*(f)$ for some f in F_X.

$$\begin{tikzcd}
\bullet \arrow[d, shift right, "p_1^*(g)"'] \arrow[d, shift left, "p_2^*(g)"] \arrow[d, phantom, "="] & u^*(\xi) \arrow[d, "g"'] \arrow[r, phantom, ""] & \xi \arrow[d, "f"] \\
\bullet & u^*(\eta) & \eta
\end{tikzcd}
\qquad g \rightsquigarrow f$$

$$\begin{tikzcd}
X' \times_X X' \arrow[r, shift left, "p_1"] \arrow[r, shift right, "p_2"'] & X' \arrow[r, "u"] & X \\
{[(U_{ij})_{i,j}]} & {[(U_i)_i]} & {[U]}
\end{tikzcd}$$

A descent datum is a family $\xi' \in F_{X'}$ of objects whose respective restrictions in $F_{X' \times_X X'}$ all agree (up to isomorphism). The condition (B) says that, in such a case, the objects in the family can actually be glued together along u, so as to be essentially equivalent to an object in F_X: $\xi' \simeq u^*(\xi)$ for some $\xi \in F_X$.

$$
\begin{tikzcd}
p_1^*(\xi') \simeq p_2^*(\xi') & \xi' \arrow[r, squiggly] & \xi \\
{[(f_i|U_{ij} \simeq f_j|U_{ij})]} & {[(f_i)_i]} & {[f]} \\
X' \times_X X' \arrow[r, shift left, "p_1"] \arrow[r, shift right, "p_2"'] & X' \arrow[r, "u"] & X \\
{[(U_{ij})_{i,j}]} & {[(U_i)_i]} & {[U]}
\end{tikzcd}
$$

Definition 6 An arrow $u : X' \to X$ in a category C with pullbacks is a *strict descent morphism* if it is a strict p-descent morphism in the particular case where p is the codomain functor $\mathrm{cod} : C^{\to} \to C$.

8.3 Algebraic Set Theory and Descent Theory

8.3.1 A Classical Result

The link of the first axioms of AST with descent theory can now be explained. Throughout this section, C will be assumed to be a category with all finite limits.

Definition 7 A *strict epimorphism* in C is the joint coequalizer of all pairs of arrows which it coequalizes.

An arrow $u : X' \to X$ is a *universal strict epimorphism* in C iff, for any $\xi : Y \to X$ in C, the pullback $\xi^*(u) : Y \times_X X' \to Y$ of u along ξ is a strict epimorphism in C/Y.

Proposition 2 (Grothendieck (1962), Proposition 2.1) *The universal strict epimorphisms in C are exactly the descent morphisms in C.*

The proof of this Proposition, that is not given by Grothendieck, requires a few basic lemmas.

Lemma 3 *Let $u : X' \to X$ be an arrow in C. Then u is a coequalizer of the pair $\langle p_1, p_2 : X' \times_X X' \to X' \rangle$ iff, for any object U of C, $Hom_C(X, U) \xrightarrow{(-)\circ u} Hom_C(X', U)$ is (up to isomorphism) the equalizer*

$$Eq(\ Hom_C(X', U) \underset{(-)\circ p_2}{\overset{(-)\circ p_1}{\rightrightarrows}} Hom_C(X' \times_X X', U)\ )$$

of the pair $\langle (-) \circ p_1, (-) \circ p_2 \rangle$.

Proof The first half of the Lemma follows from the classical fact that, for any object U of C, the corresponding representable functor $\mathrm{Hom}_C(-, U)$ sends colimits to limits, so that if u is a coequalizer of $\langle p_1, p_2\rangle$, then $\mathrm{Hom}_C(u, U) = (-)\circ u : \mathrm{Hom}_C(X, U) \to \mathrm{Hom}_C(X', U)$ is an equalizer of $\langle \mathrm{Hom}_C(p_1, U), \mathrm{Hom}_C(p_2, U)\rangle = \langle (-)\circ p_1, (-)\circ p_2\rangle$. The second half of the Lemma follows from the equally classical fact that the Yoneda embedding $y' : C^{\mathrm{op}} \to \mathit{Sets}^C$ reflects limits: If, for any U, $y'(u)(U) = (-)\circ u : \mathrm{Hom}_C(X, U) \to \mathrm{Hom}_C(X', U)$ is an equalizer of $\langle y'(p_1)(U), y'(p_2)(U)\rangle = \langle (-)\circ p_1, (-)\circ p_2\rangle$ in C, then $y'(u) : \mathrm{Hom}_C(X, -) \to \mathrm{Hom}_C(X', -)$ is an equalizer of $\langle y'(p_1), y'(p_2)\rangle$ in Sets^C (because limits are computed pointwise), so u is an equalizer of $\langle p_1, p_2\rangle$ in C^{op} and thus a coequalizer of $\langle p_1, p_2\rangle$ in C. □

Lemma 4 *Let F be a fibered category over C, $u : X' \to X$ an arrow in C, Y and Z two objects above X (i.e., two arrows $\xi : Y \to X$ and $\eta : Z \to X$ in C). Let's write $Y' := Y \times_X X'$, $Z' := Z \times_X X'$, $X'' := X' \times_X X'$, $Y'' := Y' \times_{X'} X'' = Y' \times_Y Y'$ and $Z'' := Z' \times_{X'} X'' = Z' \times_Z Z'$. One has:*
$Y' \times_{X'} Z' \simeq Y' \times_Y (Y \times_X Z)$
$Y'' \times_{X''} Z'' \simeq Y'' \times_Y (Y \times_X Z)$.

Proof

$$\begin{aligned}
Y' \times_{X'} Z' &\simeq (Y \times_X X') \times_{X'} (Z \times_X X') \\
&\simeq (Y \times_X Z) \times_X X' \\
&\simeq (Y \times_X Z) \times_Y (Y \times_X X') \\
&\simeq (Y \times_X Z) \times_Y Y' \\
&\simeq Y' \times_Y (Y \times_X Z)\ .
\end{aligned}$$

The proof about $Y'' \times_{X''} Z''$ is strictly analogous. □

Lemma 5 *Given a pullback square in C of the form*

$$\begin{tikzcd}
Y' \times_{X'} Z' \arrow[r, "p_2'"] \arrow[d] \arrow[dr, phantom, "\lrcorner", very near start] & Z' \arrow[d] \\
Y' \arrow[r] \arrow[ur, dashed] \arrow[u, dashed, bend left] & X'
\end{tikzcd},$$

one has: $\mathrm{Hom}_{C/X'}(Y', Z') \simeq \mathrm{Hom}_{C/Y'}(Y', Y' \times_{X'} Z')$.

Proof The bijection $\varphi : \mathrm{Hom}_{C/X'}(Y', Z') \to \mathrm{Hom}_{C/Y'}(Y', Y' \times_{X'} Z')$ is given by: $\varphi(f) = 1_{Y'} \times_{X'} f$ for each $f : Y' \to Z'$, and $\varphi^{-1}(g) = p_2' \circ g$ for each $g : Y' \to Y' \times_{X'} Z'$. □

Proof of Proposition 2 An arrow $u : X' \to X$ in C is a universal strict epimorphism iff, for any $Y \xrightarrow{\xi} X$, $\xi^*(u) : Y' \to Y$ is a strict epimorphism in C/Y (by Definition 7) iff, for any $Y \xrightarrow{\xi} X$ and any U above Y,

$$\mathrm{Hom}_{C/Y}(Y, U) \simeq \mathrm{Eq}(\ \mathrm{Hom}_{C/Y}(Y', U) \underset{(-)\circ\pi_2}{\overset{(-)\circ\pi_1}{\rightrightarrows}} \mathrm{Hom}_{C/Y}(Y'', U)\ ) \text{ (by Lemma 3)}$$

(where π_1 and π_2 are the two canonical projections $Y'' \to Y'$)

iff, for any $Y \xrightarrow{\xi} X$ and any $Z \xrightarrow{\eta} X$,

$$\mathrm{Hom}_{C/Y}(Y, Y \times_X Z) \simeq \mathrm{Eq}\ (\mathrm{Hom}_{C/Y}(Y', Y \times_X Z) \rightrightarrows \mathrm{Hom}_{C/Y}(Y'', Y \times_X Z))$$

iff, for any $Y \xrightarrow{\xi} X$ and any $Z \xrightarrow{\eta} X$,

$$\mathrm{Hom}_{C/Y}(Y, Y \times_X Z) \simeq \mathrm{Eq}\ (\mathrm{Hom}_{C/Y'}(Y', Y' \times_Y (Y \times_X Z)) \rightrightarrows$$
$$\mathrm{Hom}_{C/Y''}(Y'', Y'' \times_Y (Y \times_X Z)))\ \ \text{(by Lemma 5, applied twice)}$$

iff, for any $Y \xrightarrow{\xi} X$ and any $Z \xrightarrow{\eta} X$,

$$\mathrm{Hom}_{C/Y}(Y, Y \times_X Z) \simeq \mathrm{Eq}\ (\mathrm{Hom}_{C/Y'}(Y', Y' \times_{X'} Z') \rightrightarrows$$
$$\mathrm{Hom}_{C/Y''}(Y'', Y'' \times_{X''} Z''))\ \ \text{(by Lemma 4, applied twice)}$$

iff, for any $Y \xrightarrow{\xi} X$ and any $Z \xrightarrow{\eta} X$,

$$\mathrm{Hom}_{C/X}(Y, Z) \simeq \mathrm{Eq}(\ \mathrm{Hom}_{C/X'}(Y', Z') \rightrightarrows \mathrm{Hom}_{C/X''}(Y'', Z'')\ )$$

(by Lemma 5, applied three times)

iff, for any $Y \xrightarrow{\xi} X$ and any $Z \xrightarrow{\eta} X$,

$$\mathrm{Hom}_{C/X}(\xi, \eta) \xrightarrow{u^*} \mathrm{Hom}_{C/X'}(u^*(\xi), u^*(\eta)) \underset{p_2^*}{\overset{p_1^*}{\rightrightarrows}}$$
$$\underset{p_2^*}{\overset{p_1^*}{\rightrightarrows}} \mathrm{Hom}_{C/X''}((u \circ p_1)^*(\xi), (u \circ p_2)^*(\eta))\ \ \text{is exact}$$

(where p_1 and p_2 are the two canonical projections $X'' \to X'$)
iff u is a descent morphism in C (by Definition 5, condition (A)). □

8.3.2 *The Descent Condition on a Class of Small Maps*

The preceding result makes no mention of any particular class of arrows in C. Still, the third axiom of AST has to do with it.

Lemma 6 *In a Heyting pretopos C, every universal strict epimorphism is a strict (a.k.a. effective) cod_S-descent morphism, for any collection S of arrows in C satisfying the first three axioms of* AST.

Proof Let C be a Heyting pretopos and S a collection of arrows in C which satisfies the first three axioms of AST. By Lemma 1, the functor $\mathrm{cod}_S : S \to C$ is not only a fibered category, but actually a fibered subcategory of the codomain functor cod $: C^{\to} \to C$. Thus, a direct transposition of the proof of Proposition 2 shows that every universal strict epimorphisms in C is a cod_S-descent morphisms in C. So it remains only to prove that all cod_S-descent morphisms in C are strict, i.e., that the condition (B) is satisfied as well. But the proof given by Joyal and Moerdijk[15] that, in any pretopos C, every epimorphism is a strict descent morphism can be adapted here: It pertains to cod $: C^{\to} \to C$, but remains true about $\mathrm{cod}_S : S \to C$. Indeed, given any epimorphism $f : Y \to X$ in C and any $p : E \to Y$ in S above Y (i.e., $p \in S_Y$), Joyal and Merdijk show that there exists a map $r \in C/X$ and an isomorphism $s : Y \times_X Q \xrightarrow{\simeq} E$ such that:

$$\begin{tikzcd}
Y \times_X Q \arrow[r, "\simeq", "s"'] \arrow[d, "f^*(r)"'] & E \arrow[r, "q"] \arrow[d, "p"] & Q \arrow[d, dashed, "r"] \\
Y \arrow[r, "="] & Y \arrow[r, two heads, "f"'] & X
\end{tikzcd}$$

commutes. But s is an isomorphism, thus in S (by Axiom 1) and p, by hypothesis, is in S as well, so $f^*(r) = p \circ s$ is in S (since S is closed under composition, by Axiom 1 again). Then Axiom 3 exactly guarantees that r is in S, and p is isomorphic to $f^*(r)$ since S is an isomorphism over Y. □

Lemma 7 *Let S be a collection of arrows in C satisfying the first two axioms of* AST, *and let $u : X' \to X$ be an arrow in C. If u is a strict cod_S-descent morphism, then u is an epimorphism in C.*

Proof See Johnstone (2002), vol. 1, Part B, Proposition 1.5.6 (i), pp. 298–299. □

Lemma 8 *In a Heyting pretopos, every epimorphism is a universal strict epimorphism.*

Proof Indeed, any epimorphism in a pretopos is regular, and thus strict.[16] Besides, the fourth and last axiom of a pretopos, as it is defined by Joyal and Moerdijk,[17] exactly means that any epimorphism in a pretopos is universal, and thus a universal strict epimorphism. □

Joyal and Moerdijk's "Descent" Axiom, whose role has now been pinpointed, can finally be reinterpreted in terms of descent theory as follows.

Theorem 9 *Let C be a Heyting pretopos. A full subcategory S of $C^{\to}$ verifies the first three axioms of* AST *iff $\mathrm{cod}_S : S \to C$ is a fibered category and satisfies the analog of Grothendieck's result (Proposition 2), i.e., iff $\mathrm{cod}_S : S \to C$ is a fibered category and the epimorphisms in C are exactly the strict cod_S-descent morphisms in C.*

---

[15] Joyal and Moerdijk (1995), Appendix C, pp. 114–115.

[16] See Johnstone (2002), vol. 1, Part A, Corollary 1.4.9, p. 38.

[17] See Joyal and Moerdijk (1995), Appendix B, p. 109.

Proof Suppose that cod_S is a fibered category and that the epimorphisms in C are the strict cod_S-descent morphisms in C. Then the Axioms 1, 2 and 3 of AST are obviously verified. Conversely, suppose that S verifies these axioms. Then, by Lemma 1, cod_S is a fibered category. Moreover, by Lemma 8, every epimorphism in C is a universal strict epimorphism in C and thus, by Lemma 6, a strict cod_S-descent morphism in C. But, by Lemma 7, every strict cod_S-descent morphism in C is also an epimorphism in C. □

The rationale of the first three axioms of a class of small maps in AST now becomes apparent: *A class of small maps (viewed as a full subcategory of $C^{\rightarrow}$) can be partially, yet crucially, characterized as a fibered category to which descent theory applies.*

Conclusion

One should be able now to see how AST embodies a very nice combination of set theory and category theory. From a set-theoretic viewpoint, maps are in fact sets: There is no arrow. The only arrows are the edges of the graph of the membership relation: $a \in b$ means that (a, b) is an edge of the graph of $\in$, and no genuine arrow exists besides such edges, since a map is a vertex (an object) as much as its source and its target are. The algebraic set-theoretic viewpoint, on the contrary, consists in putting arrows to the fore, through the "codomain functor"—which is typical of the category-theoretic perspective. The theory of fibered categories thus reinterprets set theory so as to turn it into an arrow-based theory, rather than an object-based theory. That is why it can be described as a graft of category theory onto ZFC, which is much more fruitful than the standard rivalry between set theory and category theory.

There is much more to it. Indeed, AST accomplishes a categorical implementation of ZFC with a deep geometric twist, linked to the geometric background of the theory of fibered categories and descent. Descent theory has to do with the way in which local data can be glued together so as to produce a global item. The glueing conditions that define a strict descent morphism directly come from sheaf theory, actually—a key area of modern algebraic geometry. This paper set out to show that the main purpose of the first axioms of AST was precisely to set up a categorical reformulation of ZFC in the framework of descent theory. This perspective can actually be made very precise: A collection S of arrows in a Heyting pretopos C (a candidate for a class of small maps) verifies the first three axioms of AST iff the restriction of the codomain functor to S, $\mathrm{cod}_S : S \to C$, is a fibered category in which descent works as in the full codomain functor $\mathrm{cod} : C^{\rightarrow} \to C$; otherwise put, a fibered category in which whatever is expected to glue actually glues.

This parallel opens up a whole avenue of interactions. In addition to a remarkable example of combination between set theory and category theory—that was the starting point of this paper—, it shows a very rich interplay between set theory, on the one hand, abstract algebraic geometry and algebraic topology on the other. For instance, the additional "Representability Axiom" of AST[18] refers to the theory of classifying spaces and universal fiber bundles in topology, while introducing at the same time

[18] Joyal and Moerdijk (1995), p. 9.

an original representation of the universal class of all sets (or of all small maps). A genuine conceptual fusion is thereby carried out, which mixes foundational theories and theories apparently linked to leaves of the tree of mathematics. As a result, the distinction between “foundations of mathematics” and “mathematical practice” is seriously blurred.

References

Awodey, S., Butz, C., Simpson, A., & Streicher, T. (2014). Relating first-order set theories, toposes and categories of classes. *Annals of Pure and Applied Logic*, *165*(2), 428–502.

Freyd, P. (2004). Homotopy is not concrete. *Reprints in Theory and Applications of Categories*, *6*:1–10. Originally published in 1970.

Grothendieck, A. (1962). Technique de descente et théorèmes d’existence en géométrie algébrique I. Généralités. Descente par morphismes fidèlement plats. In A. Grothendieck (Eds.), *Fondements de la Géométrie Algébrique. Extraits du Séminaire* BOURBAKI *1957–1962*, Vol. 12 (1959/60) of *Séminaire Bourbaki*, pp. 190.01–190.29. Paris: Secrétariat mathématique.

Jacobs, B. (1999). Categorical logic and type theory. In *Studies in logic and the foundations of mathematics* (Vol. 141). Amsterdam: Elsevier.

Johnstone, P. T. (2002). Sketches of an Elephant: A topos theory compendium. In *Oxford logic guides* (Vol. 43–44). Oxford: Clarendon.

Joyal, A., & Moerdijk, I. (1995). Algebraic set theory. In *London mathematical society lecture note series* (Vol. 220). Cambridge: Cambridge University Press.

Leinster, T. (2012). Rethinking set theory. arXiv:1212.6543v1.

MacLane, S. (1997). *Categories for the working mathematician* (2nd ed., Vol. 5). Graduate texts in mathematics. Berlin: Springer.

MacLane, S., & Moerdijk, I. (1992). *Sheaves in geometry and logic*. New York: Springer.

Vistoli, A. (2005). Grothendieck topologies, fibered categories and descent theory. In B. Fantechi, et al. (Eds.), *Fundamental algebraic geometry. Grothendieck’s FGA explained. Mathematical surveys and monographs* (Chap. 1, Vol. 123, pp. 1–104). Providence: American Mathematical Society.

Chapter 9
True V or Not True V, That Is the Question

Gianluigi Oliveri

Abstract In this paper we intend to argue that: (1) the question 'True V or not True V' is central to both the philosophical and mathematical investigations of the foundations of mathematics; (2) when posed within a framework in which set theory is seen as a science of objects, the question 'True V or not True V' generates a dilemma each horn of which turns out to be unacceptable; (3) a plausible way out of the dilemma mentioned at (2) is provided by an approach to set theory according to which this is considered to be a science of structures.

Keywords Mathematical structuralism · Foundations of mathematics · Universes of sets

9.1 Introduction

One of the central problems in the philosophy of mathematics is that of determining whether mathematics is about discovering and describing the properties of existing entities such as numbers, groups, vector spaces, topological spaces, etc. or if it, rather, consists in inventing and developing complex formal games and/or constructions.[1]

[1]The former position on the nature of mathematics is known as metaphysical realism whereas the latter as metaphysical anti-realism. However, in the literature, besides being there a debate between metaphysical realists and anti-realists, there also is a much more recent controversy which sees realists about truth opposing anti-realists about truth. For a realist about truth, the truth of a mathematical statement S transcends verification, whereas for the anti-realist a mathematical statement S is true if and only if we can produce a constructive proof of S. In other words, for the anti-realist about truth, the concept of mathematical truth collapses onto that of constructive provability. The realism/anti-realism debate concerning mathematical truth was introduced by Michael Dummett (see on this Dummett 1975, 1991). For a discussion of the relationship between the metaphysical realism/anti-realism debate and the realism/anti-realism debate about truth see Oliveri (2007), Chap. 1, Sects. 7–10.

G. Oliveri (✉)
ICAR CNR, Rende, CS, Italy
e-mail: gianluigi.oliveri@unipa.it

G. Oliveri
University of Palermo, Palermo, Italy

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_9

The question above has been at the heart of the so-called ‘realism/anti-realism debate’ in the philosophy of mathematics. This is a debate which has always accompanied philosophical speculation on the nature of mathematics producing a wide spectrum of positions ranging from an overcrowded Meinongianism, according to which there is a world of objects corresponding to every consistent mathematical theory,[2] to the barren emptiness of strict formalism for which mathematical theories are just games entirely based on convention.[3]

The traditional realism/anti-realism debate received a further twist with the introduction of set theory. In fact, set theory, originally elaborated by Cantor to extend mathematical investigation beyond the realm of the finite into what he called the ‘transfinite’,[4] was subsequently shown to have a very special foundational rôle as a consequence of the discovery that mathematical theories can be reduced to/expressed within it.

The reducibility/expressibility result mentioned above is very important for the realism/anti-realism dispute in the philosophy of mathematics. For, if mathematics is about sets and their properties, the realism/anti-realism dispute on mathematics can be reformulated as a controversy concerning whether set theory discovers and describes properties of existing objects/sets or if it is, rather, a collection of complex set-theoretical games and/or constructions. Of course, engaging in this dispute is equivalent to trying to decide whether or not there is such a thing as ‘the’ universe of

---

[2] See on this Balaguer (1998).

[3] Strict formalism is the position expressed by Wittgenstein in his remarks on following a rule in the *Philosophical Investigations*. For the strict formalist, not only mathematical theories are seen as formal games, but logic itself loses, as it were, the hardness of its ‘must’. See on this Wittgenstein (1983), Sects. 185–242.

[4] The translation from the original is mine:

> …we are here obliged to make a fundamental distinction as we differentiate between:
>
> $\mathbf{II}^a$ Increasable Actual Infinity or *Transfinitum*.
>
> $\mathbf{II}^b$ Non-increasable Actual Infinity or *Absolutum*.
>
> The three examples of actual infinity previously mentioned [two of which are the positive integers, and the points lying on a circle] belong all to the class $\mathbf{II}^a$ of the Transfinite. In the same way belongs to this class the smallest supra-finite ordinal number, which I call ω; then this can be augmented or increased to the next larger ordinal number $\omega + 1$, this again to $\omega + 2$ and so on. But even the smallest actual-infinite power or cardinal number is a transfinite, and the same holds of the next larger cardinal number and so on.
>
> The Transfinite with its wealth of formations and forms points necessarily at an *Absolute*, at the ‘true infinity’, whose size can in no way be added to or decreased and which therefore, as to quantity, is to be considered as the absolute maximum. The latter so to speak goes beyond the human power of comprehension and eludes in particular mathematical determination; whereas the transfinite not only fills the wide domain of possibilities concerning the knowledge of God, but also offers a rich and always growing field of ideal research …(Cantor 1962, pp. 405–406).

sets V.[5] Indeed, believing in a true V is equivalent to being realist about set theory, because if there is a true V then the task of set theory is that of discovering and describing properties of existing objects/sets, namely, properties of the entities belonging to V; vice versa if set theory consists in discovering and describing properties of existing objects/sets then the collection of such entities is the true V.

All those who have doubts about the legitimacy, on the part of the realist, of believing in the existence of the collection of all existing objects/sets should note that since, for the realist, the objects/sets whose properties set theory investigates all exist they must also be compossible, that is, given any two existing objects/sets **a** and **b** whose properties are investigated by set theory, the existence of **a** must be compatible with the existence of **b**. From this we have that the realist is justified in believing in the existence of V, because the collection of the objects/sets whose properties set theory investigates (V) is *de facto* a coherent totality. Obviously, this does not mean that the realist is *de facto* justified in believing in the coherence of particular representations of V as, for some, we have, for instance, in the so-called 'models' of first-order ZFC.

Although several important mathematical theories have gone through a stage of development in which they were not axiomatized, and in spite of the importance of certain results, such as Gödel's First Incompleteness theorem, which state very clear limitations to what has become known as the 'axiomatic method', it is undeniable that the process of axiomatization of a mathematical theory plays a central rôle in contemporary mathematics.[6] This is especially the case when we come to theories which deal with the foundations of mathematics where rigour, and the transparency regarding their basic assumptions, are of paramount importance to prevent paradox and ensure stability to mathematics as a whole. And, since it is set theory that we are going to deal with in this paper, the considerations above should provide a convincing explanation of the reason why in what follows, instead of talking about set theory in very general terms, we shall always use the present day mainstream axiomatic system for set theory: first-order ZFC.

As is well known, if first-order ZFC is consistent a question that naturally arises is whether or not there is such a thing as 'the' universe of sets, V, that first-order ZFC is intended to describe. And, as a matter of fact, several set theorists not only believe in the existence of a true V, but identify such a V with a model of first-order ZFC known as the von Neumann cumulative hierarchy.[7] But, unfortunately, this view runs against a serious difficulty given by the fact that, if by 'universe of sets' we mean a model

[5]The symbol 'V' was introduced by G. Peano in Sect. IV of Peano (1889).

[6]There are some philosophers of mathematics, like Carlo Cellucci, who have a very negative view of the effects of the implementation of the axiomatic method within mathematical theories. For a discussion of this and related issues see Cellucci (2000, 2005), Oliveri (2005), Rav (2008).

[7]The von Neumann universe V is defined by the following equations, here *On* is the proper class of ordinals (proper classes are nor sets):

$$V_0 = \emptyset \tag{9.1}$$

$$V_{\alpha+1} = \mathcal{P}(V_\alpha) \tag{9.2}$$

of first-order ZFC, it is possible to show that if first-order ZFC is consistent it must then have a plurality of models some of which are not even going to be isomorphic to each other; and, indeed, if this is the case, how are we going to show that, within the plurality of models of first-order ZFC, it is the von Neumann cumulative hierarchy that is the intended one, that is, the true V?

The problem of whether or not there is an intended model of first-order ZFC is philosophically very important, because not only it bears on the realism/anti-realism debate in the philosophy of mathematics, on the truth of set theoretical statements, and on the epistemology and the heuristics of set theory, but it also is of crucial importance to whether or not it is possible to provide a set-theoretical foundation for mathematics.

With regard to the latter point, consider that if there were a true V we then could say that the type of entities (sets) belonging to V correctly represent the foundations of mathematics in the sense that these would be the genuinely ultimate objects from which it is possible to reconstruct any entity studied by a mathematical theory reducible to/expressible within first-order ZFC. But, on the other hand, if there were no true V, the entire problem of providing a foundation for mathematics in terms of sets from which it is possible to reconstruct any entity studied by a mathematical theory reducible to/expressible within first-order ZFC would appear to lack sense as a consequence of the non-uniqueness of such a reconstruction.

To this someone like Paul Benacerraf might reply that even if we have a true V there is no guarantee that we must have a unique set-theoretical reconstruction of mathematical entities as it happens, for example, in the case of the natural numbers.[8] But, even though this is certainly true and independently of the classical structuralist reply to Benacerraf's observation,[9] we must point out that the phenomenon described

(Footnote 7 continued)

$$V_\beta = \bigcup_{\alpha<\beta} V_\alpha \quad \text{if } \beta \text{ is a limit ordinal} \tag{9.3}$$

$$V = \bigcup_{\alpha\in On} V_\alpha. \tag{9.4}$$

[8]Indeed, if by ‘Zermelo sequence’, (Z_n), and ‘von Neumann sequence’, (N_n), we mean the sequences of sets individuated by the following pairs of recursive equations:

$$Z_0 = \emptyset \text{ and } Z_{n+1} = \{Z_n\};$$

$$N_0 = \emptyset \text{ and } N_{n+1} = N_n \cup \{N_n\};$$

it is well known that both (Z_n) and (N_n) are adequate set-theoretical representations of the sequence of natural numbers. Of course, the embarrassing thing here is given by the observation that (Z_n) and (N_n) give different set-theoretical representations from one another of all the natural numbers $n \geqslant 2$. Benacerraf discusses some of the philosophical implications of the existence of (Z_n) and (N_n) in Benacerraf (1985).

[9]The classical structuralist reply to Benacerraf's problem—What is the correct set-theoretical representation of the natural numbers?—is to say that the apparent impossibility of solving Benacerraf's problem, as a consequence of the existence of the sequences (Z_n) and (N_n) of footnote 8, is not important, because what is mathematically relevant to number theory is the number theoretical structure (see on this Sect. 9.5 especially footnote 35), and from Dedekind's recursion theorem

by Benacerraf does not undermine the foundational rôle of V, because if there is a true V then regardless of how you want to represent, say, the number 2 in V—as $\{\{\emptyset\}\}$ or as $\{\emptyset, \{\emptyset\}\}$ or as some other set belonging to some other adequate sequence of sets—you, nevertheless, produce your representations of 2 by means of entities all of which belong to V, that is, by means of V-sets.[10] Therefore, whichever adequate representation of the natural numbers (within V) you choose, and in particular of the number 2, V-sets would always have to be considered as 'the basic furniture of arithmetic'. It is interesting to notice that a situation similar to that we have described above obtains in physics. In quantum physics, although you can adequately represent the electromagnetic radiation by means of either discrete particles (photons) or continuous waves, both particles and waves belong to the same universe. On the other hand, if we had two different set-theoretical universes V_1 and V_2, the foundational rôle of both V_1 and V_2 would be irremediably compromised, because if the V_1-sets used to represent the natural numbers were not V_2-sets, and vice versa, we would be unable to say what kind of things the natural numbers are, that is, whether they are V_1-sets or V_2-sets.

'True V or not true V' is a very important question also from a mathematical point of view, because if there were a true V then: (1) the truth or falsity of the continuum hypothesis (CH), and of any other set-theoretical statement independent of first-order ZFC, would be a genuine mathematical open question which in principle could be sorted out individuating the relevant features of V and representing them by means of supplementary axioms to be added to first-order ZFC; (2) it would be possible to provide a very natural explanation of the otherwise rather mysterious process whereby the axioms of first-order ZFC have been chosen: they would be extrapolations/abstractions of important characteristics of V which enable us to prove theorems about V and its elements.

But, if there were no true V then any set-theoretical statement S independent of first-order ZFC could not count as an open mathematical question, because, apart from the absence of a procedure of decision for S within first-order ZFC, there would be no fact of the matter related to it. Moreover, if statements such as S cannot be considered as open mathematical questions, we should then either 'bite the bullet' and say that these are not mathematical statements at all or, if we are not prepared to do that, we should eliminate the law of excluded middle from our system of logic for set theory, because statements such as S clearly lack a truth value, and treat all statements that are independent of first-order ZFC as ideal statements in Hilbert's sense.[11]

(Footnote 9 continued)
we know that in second-order arithmetic any two models of Peano axioms are isomorphic to one another, i.e. they have the same structure.

[10]Note that the lowest type of the von Neumann universe where $\{\{\emptyset\}\}$ and $\{\emptyset, \{\emptyset\}\}$ occur as elements is V_3.

[11]Given a mathematical theory T, Hilbert distinguishes between *finitary* and *ideal* statements belonging to the language of T. If T is number theory then a finitary number theoretical statement is a statement the truth of which can be established by a finitely long computation, e.g., $2 + 3 = 5$. Note that the class of finitary number theoretical statements $\mathfrak{F}$ is closed under finitely many applications of

In this paper we intend to argue that: (1) the question 'True V or not True V' is central to both the philosophical and mathematical investigations of the foundations of mathematics; (2) when posed within a framework in which set theory is seen as a science of objects, the question 'True V or not True V' generates a dilemma each horn of which turns out to be unacceptable; (3) a plausible way out of the dilemma mentioned at (2) is provided by an approach to set theory according to which this is considered to be a science of structures.

Note that the view of set theory foreshadowed in (3) above motivates an approach to formal systems of set theory in general according to which such systems aim at describing/investigating the structures realized by their models. This, besides being similar to what ordinarily happens in group theory and in topology, raises an important question which is about what happens within such a structuralist framework for set theory to all that work that is labelled as 'foundations of mathematics'.

In Sect. 9.2 we will discuss some of the issues surrounding the problems of what we ought to mean by 'acceptable foundation of mathematics', and of whether first-order ZFC can provide an acceptable foundation of mathematics.

In Sect. 9.3 we will examine aspects of the general framework within which the central question of the paper can be formulated. We are going to argue that in such a framework the concepts of faith and persuasion appear to play a non-secondary rôle in set theory blurring somewhat the traditional demarcation line between proof and rhetorical argument, and revealing the profoundly conjectural nature of mathematical theories.

In Sect. 9.4 we are going to formulate the dilemma generated by the central question of the paper when this is expressed in a context where set theory is considered to be a 'science of objects'. We will also discuss some of the main philosophical and mathematical consequences of the dilemma.

In the fifth and final section we will endeavour to show that the dilemma can be resolved within a framework according to which set theory is a science of structures, and will explore some of the consequences of adopting the structuralist view of set theory here presented on the debate concerning the foundations of mathematics.

9.2 First-Order ZFC and the Foundations of Mathematics

In the introduction we formulated the main question of the paper—true V or not true V—in relation to the axiom system of set theory known as first-order ZFC. Moreover, we added that, in case first-order ZFC is consistent, the question whether there exists an intended model of first-order ZFC has important consequences for the possibility

(Footnote 11 continued)
$\neg, \wedge, \vee$. From this it easily follows that a clear example of ideal number theoretical statement is: for any $m, n \in \mathbb{N}$, $m + n = n + m$. According to Hilbert, although ideal statements 'signify nothing' (Hilbert 1926, p. 196), they are useful to develop mathematical theories, because they contribute to keeping the laws of logic and mathematics in their simplest form. See on this Hilbert (1926), p. 195.

of finding a set-theoretical foundation of mathematics. Now, before proceeding any further, we need to give an explanation of the special rôle attributed in this paper to first-order ZFC, and of the envisaged connection between the existence of an intended model of first-order ZFC and the foundations of mathematics. But, in order to do so, let us first address the vexed question concerning what a foundation of mathematics ought to be.

The term ‘foundations’, occurring in the phrase ‘foundations of mathematics’, clearly appeals to an architectonic metaphor according to which mathematics, like a building, is erected on a base-level theory which provides a safeguard against threats to its stability.[12] Historically the function of base-level theory for mathematics has been taken in turns by logic, set theory and, more recently, by category theory, whereas the threats against the stability of mathematics have been so far identified with: (1) the obscurity of basic notions such as those of natural number and set; (2) problems relating to the concept of mathematical justification/proof; (3) the logical, semantical, and set-theoretical paradoxes discovered during that remarkable period of time going from the end of the XIXth century to, approximatively, the first half of the XXth century.[13]

Although it is clear how the appearance of paradoxes within a mathematical theory T which makes use of classical logic can be a threat for the stability of T—the paradoxes trivialize T—it is not so obvious what threat might be posed to the stability of T by problems surrounding either the nature of its basic entities, or the concept of proof in T. To see this consider, first, that if numbers are thought to be objects it then makes little sense to believe that we have a theory of numbers unless we know what kind of objects numbers are, and can provide a convincing account of how we come to know that they exist and have certain properties. These were the kind of difficulties, present in late XIXth century philosophy of mathematics, that started Frege thinking about the foundations of arithmetic[14] and, eventually, led him to his peculiar version of logicism.[15]

Secondly, if we consider arithmetic as the theory within which we find and justify truths about numbers, it is important we pay some attention to the process whereby we establish such truths. For, if the elements of the set of assumptions/axioms from which we start in producing our proofs are not self-evidently true nor is such a set of assumptions/axioms provably consistent, it follows that the deductive chains by us produced within arithmetic cannot be taken to establish the truth of the statements deduced, but only that these are logical consequences of the assumptions/axioms used in their deduction.

[12] See on this Ferreiros and Gray (2006).

[13] The first set-theoretical paradox to be published was Burali-Forti’s in 1897 (see Burali-Forti 1967a, b), whereas one of the latest to appear in print was Montague’s paradox of groundedness in 1955 (see Montague 1955).

[14] See on this Frege (1980).

[15] For Frege, there is no demarcation line between arithmetic and logic, but there is a demarcation line between geometry and logic: local logicism. On the other hand, for Russell, the whole of mathematics is part of logic: global logicism.

In the presence of the challenge posed by difficulties (1)–(3) above to the way of thinking about the foundations of mathematics we have examined so far, we must conclude that these cast a long shadow on a very important idea that from the beginning has been associated with mathematics: the certainty of its methods. Given what we have just said, it is clear how the most pressing question that now needs to be addressed is 'Is it possible to establish once and for all the certainty of the methods adopted in mathematics?' We shall call this the 'First Problem of Foundations' (FPF), and it is interesting to notice that FPF was the main problem that should have been sorted out by the realization of what, within foundational studies, became known as 'Hilbert's programme'.

Since the failure of Hilbert's programme, coming at the hands of Gödel's Second Incompleteness Theorem, appears to answer the FPF with a clear and resounding 'No', if we put constructivism to one side, it seems there is no way that the main task placed before the follower of the tradition in the foundations of mathematics inspired by the architectonic metaphor can be accomplished.

However, besides the view of the foundations of mathematics guided by the architectonic metaphor, a view driven by an obsession with what we might call 'structural stability', there is a much older tradition which considers a work in any subject $\mathfrak{S}$ to be foundational if and only if it is concerned with the prime notions/posits from which it is possible to define/reconstruct all the entities which are object of study in $\mathfrak{S}$.

At the heart of this latter sense of 'foundational' there is a profound metaphysical concern which, on one hand, has to do with order and unification emerging from caos and, on the other, with an idea connected with the principle known as Occam's razor. This is the idea according to which, since in providing an explanation of a given phenomenon one should not multiply the possible assumptions *ad libitum* (Occam's razor principle), it follows that we have an ontological commitment only towards those entities presupposed by the best, and most ontologically economical, explanation of the phenomenon. The main problem emerging in a very natural way from this line of thought on the foundations of mathematics is: What are the prime/basic objects of mathematics? We shall call this the 'Second Problem of Foundations' (SPF). Russell's programme of reduction of mathematics to logic, and the idea of reducing mathematics to set theory, can be seen as obvious attempts to solve this problem.

As a matter of fact, this latter attitude towards the foundations of mathematics, besides being very old in its reaching back to the roots of the axiomatic method for how this was developed at the origin of Euclidean geometry, is also common to the physical sciences. Indeed, both in the suggestions made by the pre-Socratic philosophers concerning the basic principles from which originates all there is and in the more recent investigations concerning the unification of the fields of forces, we find the pervasive presence of Occam's razor and of its ontological agenda.

Even though the situation concerning the problem of what the prime/basic objects of mathematics are appears on the surface to be hopeless—in the case of set-theoretical reductions there is a multiplicity of possible prime/basic entities to choose from: well-founded collections, non-well-founded collections, not too large collections, well-founded and not too large collections, definable collections, etc.—since mathematical practice/the mathematical community seems, for various reasons, presently to have settled on a particular axiom system of set theory known as first-order ZFC, it follows that a somewhat specialized version of SPF retains a great philosophical value. Such a version of the SPF runs as follows: What are the prime/basic sets that first-order ZFC is intended to describe and from which it is possible to reconstruct any object studied by any mathematical theory T reducible to/expressible within first-order ZFC? (SPF*).

If the above last few remarks show what the connection is between the existence of an intended model of first-order ZFC and the foundations of mathematics—consider SPF*—almost nothing has been said about our choice of first-order ZFC out of a large plurality of axiomatic systems for set theory available.

The main reason for our choice is that first-order ZFC is the current mainstream axiomatic system for set theory adopted by the mathematical community. Note that this is not a point pertaining to the sociology of mathematical knowledge. For the choice of a mathematical theory operated by the mathematical community, among a plurality of rival theories, is not ultimately consequence of irrational socio-psychological factors affecting the behaviour of mathematicians, but of a critical evaluation of the theories in question, and of some of their mathematical consequences. We have written ‘ultimately’ here, because, as the study of the history of mathematics shows, mathematical knowledge does not grow within an ideological vacuum, but within a constellation of most various beliefs and intellectual fashions with which it actively interacts. And this, of course, applies not only to the process of theory choice, but also to that concerning the simple introduction of new ideas. Very well known cases of what we might call ‘ideological interference’ with the process of growth of mathematical knowledge are: (i) the sad fate that befell the first person who divulged the existence of incommensurables, existence of incommensurables which irremediably undermined one of the central tenets of the Pythagorean school: that all things are made of positive integers (or by their ratios); (ii) the blindness shown for many centuries by Western mathematicians to the concept of 0, a blindness mainly due to the conflation of the concept of nothingness with that of vacuum; (iii) the hostility against the use of actual infinity in mathematics and, in particular, against Cantor’s theory of transfinite sets and their cardinal arithmetic, hostility caused in part by the belief that infinity is an attribute of God, etc.

However, having said that the choice on the part of the mathematical community of first-order ZFC as foundational system for mathematics is ultimately consequence of a critical evaluation of technical advantages and disadvantages deriving from the adoption of first-order ZFC, we might as well mention some of these.

A positive technical feature of first-order ZFC is that this system disposes efficiently of the known set-theoretical paradoxes. And here the term ‘efficiently’ indicates that first-order ZFC not only eliminates the known set-theoretical paradoxes, but that, in contrast with other foundational systems, e.g. Russell’s Ramified Theory of Types (RTT), it does not hindrance standard mathematical activity.

Secondly, the completeness and compactness of first-order logic are a precious addition to the strength and logical perspicuity of the axiomatic system. With regard to the logical perspicuity of first-order logic note that if, as Quine says, second-order logic is just set theory in sheep's clothing, the logical character contributed by second-order (and higher order) logic(s) to the study of set theory is undermined somewhat.

Thirdly, there is a beautiful continuity between Cantor's original system of naïve set theory (NST), and first-order ZFC, which does not exist between NST and other axiomatic systems.[16]

Fourthly, the so-called 'proof of the pudding', that is, the resilience of first-order ZFC under the test of time.

Among the disadvantages of first-order ZFC, we can mention that: (a) this system admits of infinitely many axioms; (b) the representation of a mathematical theory T within first-order ZFC often obscures the surveyability of the specific domain[17]; (c) the use of first-order logic seems to run against the higher-order character of several important results present in various areas of mathematics[18]; (d) category theory is not representable within first-order ZFC.[19]

9.3 Faith and Persuasion in Set Theory

Having produced a preliminary clarification of the meaning and the philosophico-mathematical importance of the main question of the paper (Sect. 9.1), and of the foundational relevance of the axiom system known as first-order ZFC (Sect. 9.2), before proceeding any further along the lines sketched in Sect. 9.3, we need to pin firmly down what we must mean by 'universe of sets'.

The literature concerning what the term 'V' ought to refer to contains different opinions on this subject. These range from Hamkin's idea that 'V' must be used to refer to a multiverse/hyper-universe of sets, that is, to the collection of all classes of objects which realize the various possible distinct concepts of set,[20] to the view that 'V' ought to refer to a model of first-order ZFC such as, for instance, the von Neumann cumulative hierarchy.

[16] The continuity between NST and first-order ZFC is such that it is appropriate to speak of these two theories as being part of a Cantor-Zermelo mathematical research programme. See on this Oliveri (2006) and Oliveri (2007), Chap. 7.

[17] On this, and against the very idea of a set-theoretical form of reductionism, see E. De Giorgi's programme concerning the foundations of mathematics in Forti et al. (1996), pp. 1–2.

[18] Examples of second-order mathematical statements are: The Peano axiom of full induction, Every infinite subset of a countable set is countable, Every non-empty set of reals bounded from above has a least upper bound in $\mathbb{R}$, the Bolzano-Weierstrass theorem, etc. See on this also Isaacson (1987).

[19] Some categories, e.g. the category of sets **Set**, are too large to be expressed/defined in first-order ZFC, i.e. **Set** is not a set in first-order ZFC. See on this McLarty (1995), Part II, Chap. 12, Sect. 12.1, pp. 107–110.

[20] See on this Hamkins (2011).

Now, given the rather perplexing notion of multiverse/hyper-universe of sets both in Hamkin's sense and in Sy Friedman's more modest view,[21] we decided to adopt in what follows a minimalist position according to which: (1) the concept of multiverse/hyper-universe, both in Hamkin's and in Friedman's sense, cannot offer a useful contribution to the discussion of the main issue of this paper; (2) a universe is simply a model of first-order ZFC.

The reason why we find perplexing the idea of a multiverse/hyper-universe of sets, both in Hamkin's and in Friedman's sense, is that this appears to betray the foundational vocation of set theory which, as we have seen in the previous section, no longer consists in establishing once and for all the certainty of the methods adopted in mathematics, but rather in individuating the prime/basic entities, sets, from which it is possible to reconstruct the objects of study of all mathematical theories reducible to/expressible within set theory. And indeed, if anything goes in set theory, in the sense that the task of set theory is that of describing a hyper-universe in which, in some parallel sub-universes, sets are well-founded collections, whereas in others they are not; in which, in some parallel sub-universes, there are large cardinals of any size, whereas in others this is not the case; etc. it seems as if there were a radical and unresolvable equivocity within set theory about what one should mean by 'set'. And, of course, this is precisely the kind of equivocity that is bound to be fatal for the very attempt to provide a set-theoretical foundation of mathematics in the above mentioned sense.

To this someone might reply that there is nothing to worry about the equivocity of the term 'set', because this is perfectly analogous to the equivocity of, for example, the term 'triangle' in geometry where such an equivocity, rather than paralyzing mathematical research, has enriched it with a wealth of new interesting geometries proving to be a case of what philosophers of mathematics like E. Grosholz would call 'productive ambiguity'.[22] Indeed, it is well know that, although in geometry the term 'triangle' refers to a portion of the plane enclosed by three straight lines, in Euclidean geometry the sum of the internal angles of a triangle is 180° whereas this is not the case in hyperbolic and elliptic geometries.

However, we believe the analogy drawn above between geometry and set theory to be profoundly misleading. For if the task of geometry is that of studying certain properties of structures, structures such as those described by Hilbert's axioms of Euclidean geometry,[23] there is nothing wrong with the existence of a family of different geometrical structures $(\mathcal{S}_i)_{i\in I}$ such that, for any $i, j \in I$, if $i \neq j$ then the reference of the term 'triangle' in $\mathcal{S}_i$ is an entity that differs in some universal property from the entity that is the reference of the term 'triangle' in $\mathcal{S}_j$.[24] But we can go even further than that, because if the task of geometry is that of studying certain properties

[21]For Friedman, a multiverse/hyper-universe is simply a non-empty collection of more than one model of, say, first-order ZFC.

[22]See on this Grosholz (2007) and Oliveri (2011).

[23]Tuller (1967), Appendix 2, pp. 182–185.

[24]By 'universal property' we mean a property common to all, say, triangles in $\mathcal{S}_i$ such as: the sum of the internal angles of a triangle is 180°.

of structures, it does not even matter what sort of things are those to which we refer by means of really basic terms[25] like 'point', 'line', and 'plane' in so far as these things, whatever they are, satisfy the axioms of our geometrical system.

But, when from geometry we turn our attention to set theory we realize that the situation has radically changed. For, if we stay with the idea that the main foundational concern of set theory is the individuation of the prime/basic entities, sets, from which it is possible to reconstruct the objects of study of all mathematical theories reducible to/expressible within set theory, we have that an equivocity regarding the concept of set implies that: (1) we cannot know which the prime/basic entities etc. are; and that, therefore, (2) the foundational task of set theory must remain unfulfilled.

Therefore, for the sake of preserving the foundational rôle of set theory in mathematics within a general framework in which set theory is seen as a science of objects (sets), we are going to put to one side both Hamkins' and Friedman's inflationary views of the universe of set theory, and identify V with a model of first-order ZFC. But, before proceeding any further, some cautionary remarks are in order, because, in spite of the large amount of effort and ingenuity that has gone in this direction, the activity dedicated to the study of models of first-order ZFC is to a large extent purely conjectural.

Indeed, as is well known, one of the consequences of Gödel's Second Incompleteness Theorem is that, if first-order ZFC is consistent then we cannot prove that this is the case within first-order ZFC. From this we have that our talk of 'models of first-order ZFC' turns out to be entirely based on faith, that is, on the belief in the unprovable (within first-order ZFC) assumption that first-order ZFC is consistent.

As it might be expected, the faith in the consistency of first-order ZFC shown in discussions relating to models of this formal system is not blind, but relies on good reasons some of which have been listed in the previous section among the positive features of first-order ZFC. But, although such good reasons—and, in particular, that regarding the fact that first-order ZFC disposes of the known set-theoretical paradoxes—give rational support to the belief in the consistency of first-order ZFC and, therefore, succeed in persuading mathematicians that this must be the case, they nevertheless fall short of being a proof. This is an important point, because, the impossibility of a proof of the consistency of first-order ZFC (within first-order ZFC, if first-order ZFC is consistent), besides putting a severe restraint on loose talk concerning models of first-order ZFC, shows that even those mathematical theories which relate to the foundations of mathematics—theories where clarity and rigour are of paramount importance—are essentially conjectural and that the very concept of 'proof in first-order ZFC', being ultimately based on persuation and faith, is no longer distinguishable in a very sharp way from what we might call 'rhetorical argument'.

At this point it is important to notice that the rôle of faith in set theory does not stop at the belief in the consistency of first-order ZFC, but also grounds, for example, the extremely elaborate theory of large cardinals. To see this consider that in first-order

[25] Here by 'really basic terms' we mean terms belonging to the language of a mathematical theory T which are not in need of an explicit definition. Within an axiomatic mathematical theory T, the relevant part of the meaning of its really basic terms is given implicitly by some of the axioms of T.

ZFC it can be proved that if κ is an inaccessible cardinal[26] then V_κ, i.e. the initial part of the von Neumann cumulative hierarchy indexed by κ (see footnote 7), is a model of first-order ZFC.[27] This means that we can show, in first-order ZFC, that if there exists an inaccessible cardinal κ then first-order ZFC is consistent. From this it follows that if we could prove within first-order ZFC that there exists an inaccessible cardinal κ, we could then also obtain (by *Modus Ponens*) a proof in first-order ZFC that first-order ZFC is consistent. But, since this latter result would contradict Gödel's Second Incompleteness Theorem, we have that the existence of such a κ cannot be proved within first-order ZFC. The same reasoning, of course, applies to any large cardinal λ such that $\lambda > \kappa$.

In taking stock of some of the results obtained in this section, we must mention the fact that, in contrast with a very old tradition reaching back to the ancient Greeks, when we look at that part of contemporary mathematical practice present in the development of set theory, we must conclude that there is no sharp demarcation line between proof and rhetorical arguments in mathematics.

Secondly, mathematical knowledge, even that produced in some of the best foundational systems like first-order ZFC, is ultimately conjectural.

Lastly, one of the consequences of the conjectural nature of mathematical knowledge is that mathematical theories are fallible. Now, this is an extremely important result, because the fallibility of mathematical theories delivers a tremendous blow against another very old and strong prejudice about mathematics, a prejudice according to which mathematics is an intellectual endeavour that differs entirely from empirical science, because, in contrast with empirical science, mathematical knowledge grows in a purely cumulative way.

9.4 The Dilemma and Some of Its Consequences

In the previous sections we have repeatedly said that the dilemma 'True V or not true V' is generated by the fact that if first-order ZFC is consistent, and 'V' refers to a model of first-order ZFC, it can then be shown that first-order ZFC has a plurality of non-isomorphic models.[28] The source of information about the existence of such non-isomorphic models of first-order ZFC is mainly provided by three results: Gödel's First Incompleteness theorem; the Löwenheim–Skolem theorem; forcing.

Indeed, from Gödel's First Incompleteness theorem we have that, if first-order ZFC is consistent, there exist closed well-formed formulae G, belonging to the language of first-order ZFC, which are independent of first-order ZFC. If we, now,

[26]Intuitively an inaccessible cardinal κ is a cardinal greater than $\aleph_0$ that cannot be generated from smaller cardinals by means of the usual set-theoretic operations. For a more rigorous definition see Jech (2003), Part I, Chap. 5, p. 58. Note that inaccessible cardinals are the smallest large cardinals.

[27]See on this the proof of **Lemma 12.13** in Jech (2003), Part I, Chap. 12, p. 167.

[28]We overlook here the question posed by the existence of isomorphic models of first-order ZFC, because the existence of such models would neither affect ZFC truths (not just first-order truths), nor the cardinality of the domains of the models.

extend the axiomatic basis of first-order ZFC adding to it as supplementary axiom in one case G and in the other $\neg$ G, we obtain two consistent extensions of first-order ZFC whose models are not isomorphic to one another, because in a model of the first extension G is true and $\neg$ G is false, whereas in a model of the second extension G is false and $\neg$ G is true.[29]

Secondly, since first-order ZFC is a first-order theory with a countable language then, if first-order ZFC is consistent, it follows, from the Löwenheim-Skolem theorem, that first-order ZFC must have a countable model. But, of course, since the domain of the countable model of first-order ZFC cannot be put in bi-univocal correspondence either with the von Neumann cumulative hierarchy or with L,[30] we have that if first-order ZFC is consistent then there are models of first-order ZFC which are not isomorphic to one another.

Lastly, the forcing technique, introduced by Cohen in 1963, is used to individuate/construct models of first-order ZFC which are not isomorphic to one another to prove the independence of certain propositions, e.g. the Continuum Hypothesis (CH), from first-order ZFC.[31]

Having spelled out some of the technical reasons that lie at the roots of the dilemma, let us, now, go back to our discussion of its consequences. Although we have already touched upon this topic in Sect. 9.1, we need to delve a little deeper into it to show that some such consequences are truly unacceptable.

First of all, we should notice that if: (a) mathematics is a science of objects, (b) these objects are ultimately sets, and (c) there is a true V, then not only should we be realists about objects (sets), but we should also uphold a very robust realist conception of truth.

To see this latter point, we need preliminarily to consider that a conception of truth is said to be realist if, given a statement S, the truth of S transcends/is independent of the verifiability of S.[32] Having granted this much, we should now realize that: (i) if there is a true V, there must also be a fact of the matter in V determining whether a statement S belonging to the language of first-order ZFC is true or false; and that (ii) the existence of such a fact of the matter in V determining whether S is true or false has nothing to do with our ability to know how things actually are in relation to the truth or falsity of S.

But, on the other hand, if there is no true V, we would have to be committed either to anti-realism, like the formalists and the constructivists (see on this Sect. 9.1), or to some kind of Meinongianism about objects, like the followers of the so-called 'plentiful platonism' (Balaguer, see on this Sect. 9.1). It is important to notice that, in both these cases, statements S which belong to the language of first-order ZFC

[29]In other words, the models of the two extensions of first-order ZFC (and, *a fortiori*, of first-order ZFC) here mentioned are not isomorphic to one another, because they are not elementarily equivalent.

[30]With regard to L, the universe of Gödel's constructible sets, see Jech (2003), Part II, Chap. 13, pp. 175–200.

[31]See on this Jech (2003), Part II, Chap. 14, pp. 201–224.

[32]The truth of a statement S transcends the verifiability of S if and only if the statement S is true/false independently of our ability to know which is the case.

and are independent of first-order ZFC would lack a truth value, because they would neither be provable/refutable within first-order ZFC, nor would there be a fact of the matter determining whether they are true or false (see on this Sect. 9.1).

However, independently of considerations concerning if and how our beliefs about the existence of a true V should condition our choice of the type of logic (classical vs. intuitionistic, etc.) to adopt, the real problem here is that, whichever horn of the dilemma we choose, we are going to find ourselves in a very difficult situation. For, if we decide to take the true V option, we will not be able to make sense of the traditional and perfectly legitimate mathematical activity consisting in developing both the axiom systems: first-order ZFC + CH and first-order ZFC + $\neg$ CH. If we go, instead, for the not true V horn of the dilemma, we should then renounce any foundational rôle for first-order ZFC.

To see this, first of all, consider that the mathematical interest in developing rival/mutually incompatible axiom systems—like first-order ZFC + CH and first-order ZFC + $\neg$ CH—became established in the XIXth century. It was the development of the axiomatic method, taking place mainly in connection with the proof of the independence of the fifth postulate from the other postulates and axioms of Euclidean geometry, that initiated this special kind of immensely fruitful mathematical tradition which, among other things, led to the discovery of the non-Euclidean geometries.

Secondly, the activity of developing both first-order ZFC + CH and first-order ZFC + $\neg$ CH is legitimate, because, as we have already seen above, if first-order ZFC is consistent, and CH is independent of first-order ZFC, we then have that both first-order ZFC + CH and first-order ZFC + $\neg$ CH are consistent.

Thirdly, in direct contrast with the points just made above, for a believer in the existence of a true V, it is nonsensical to develop both first-order ZFC + CH and first-order ZFC + $\neg$ CH. The reason for this is that if you believe in the existence of a true V then, for you, the main task of formal systems of set theory is describing the properties of the true V and of its elements. Therefore, if in the true V CH is true, we have that $\neg$ CH must be false and, consequently, first-order ZFC + $\neg$ CH is mathematically useless in the sense that, very much like a falsified theory belonging to the empirical sciences, it does not describe the properties of the true V and of its elements.

Lastly, the situation of someone who does not believe in the existence of a true V is no better than that of the believer in a true V. For, as we have already seen in a previous section of this paper, the philosopher of mathematics for whom there is no true V should abandon any hope of using first-order ZFC to produce a Russellian foundation for mathematics. And, of course, this is bound to be unacceptable for anyone whose interest in first-order ZFC is entirely dependent on seeing this axiom system of set theory as what can provide a Russellian foundation for mathematics. At this point the pressing question becomes: where do we go from here?

9.5 All Structures Great and Small

What we said in the last section made it clear that, if we formulate the dilemma within a context characterized by what we called ‘realism about objects’ then, whichever horn of the dilemma we decide to take, the consequences of our choice will turn out to be unacceptable. This situation, of course, requires a radical rethinking of the frame of reference we have used so far and, in particular, of one of its metaphysical components represented by the idea that mathematics is a science of objects.

The reason why we decided to focus on this aspect of our frame of reference is, in part, due to the problem we came across in Sect. 9.1 (Benacerraf’s problem) which for its resolution requires a structuralist approach (see footnote 9); and, in part, to relatively recent developments of mathematics that have challenged the traditional view of mathematics as a science of objects.

Indeed, the difficulty posed by the existence of a plurality of set theoretical models of Peano arithmetic (Benacerraf’s problem), is strikingly similar to that which is at the root of the dilemma. Moreover, we know that: (1) if we want to solve Benacerraf’s problem we need to think of number theory as what describes the properties of a structure; and that (2) all those considerations relating to Benacerraf’s problem and its solution are not confined to the natural numbers, but can be generalized, for example, to the rationals and the reals which, unsurprisingly, in the literature are always defined *up to isomorphism*.

Secondly, apart from producing a survey of the type of research presently conducted in mathematics, one of the most impressive pieces of evidence in favour of the change of focus of mathematical activity from objects to structures is given by mathematics textbooks. As a matter of fact, when we open a modern textbook of algebra, we come across the definitions of group, ring, field, vector space, etc. given in terms of a domain of objects[33] concerning which one or more constants may be singled out, and on which one or more operations may be defined. Something very similar happens in topology textbooks when the definition of topological space is given,[34] and even the natural numbers, those ancient ancestors of the entire number

[33]Here by ‘object’ we simply mean an element of the domain of discourse D. For a discussion of the notion of object adopted in this paper see Oliveri (2012), Sects. 3–5.

[34]Let $(D, \mathfrak{I})$ be a topological space in which the topology $\mathfrak{I}$ is individuated by the interior operator i such that $i : \mathcal{P}(D) \to \mathcal{P}(D)$ and:

1. $i(D) = D$;
2. $i(A) \subset A$;
3. $i(A \cup B) = i(A) \cup i(B)$;
4. $i(i(A)) = i(A)$.

system, can be seen as the elements of the domain $\mathbb{N}$ of a type of structure known as 'Peano system'.[35]

Having given some justifications for the 'structuralist turn' taken by our discussion of the dilemma, we now need to provide a sharpening of the notion of structure. For this is a notion that is going to be at the heart of our concerns in the final part of this section.

Although the concept of structure we have been using so far might be intuitively clear, it is worth pointing out that a natural and more stringent characterization of it can be given in terms of an ordered pair $(\mathrm{D}, \mathfrak{R})$, where D is a non-empty domain and $\mathfrak{R}$ a set of constants and/or properties and/or functions and/or relations defined on D. To this we can add that structures can be said to be either *great* or *small*. In what follows, we are going to call a structure $(\mathrm{D}, \mathfrak{R})$ 'small' if D is a set, and 'great' if D is a proper class.[36]

Note that in the case of groups, rings, fields, vector spaces, topological spaces, Peano systems, etc. we are justified in talking about (small) structures as existing entities, because: (a) there is plenty of legitimate instantiations/realizations of such structures; (b) we have criteria of identity for these entities, criteria of identity represented by the various concepts of isomorphism specific to each of them. What we mean by (b) is that '$(\mathrm{D}_1, \mathcal{A})$' and '$(\mathrm{D}_2, \mathcal{B})$' refer to the same structure if and only if there exists a one-to-one correspondence f from D_1 to D_2 that preserves the operations or the open sets or the order, etc. A crucial feature of the identity conditions for structures is that these are independent of what sort of things the elements of D_1 and D_2 respectively are.

If we now return to set theory and, in particular, to the standard characterization of models of first-order ZFC, we realize that a (transitive) model of first-order ZFC is an ordered pair $(\mathrm{M}, \in)$ such that: (1) M is a (transitive) class,[37] i.e. M is either a (transitive) set or a (transitive) proper class, (2) $\in$ is the membership relation defined on M, and (3) all the axioms of first-order ZFC are true in $(\mathrm{M}, \in)$. But, of course, as it is clearly shown by the notation, the ordered pair $(\mathrm{M}, \in)$ squarely falls under the concept of structure we have just elucidated. And, indeed, as in the case of all the other types of structures, the identity conditions for models of first-order ZFC are precisely those given in terms of the concept of isomorphism specific to them.

If we have argued correctly so far, we have that first-order ZFC turns out to be the formal system of a theory of structures rather than objects. The consequences

---

[35] A Peano system is an ordered triple (N, o, S) such that $\mathrm{o} \in N$, $S : N \to N$ and:

(i) $\mathrm{o} \neq S(n)$, for any $n \in N$;
(ii) $S(n) = S(n') \Rightarrow n = n'$, for any $n, n' \in N$;
(iii) for any $A \subseteq N$, if $\mathrm{o} \in A$ and, for all $n \in N$, if $n \in A \Rightarrow S(n) \in A$, then $A = N$.

The natural numbers are the elements of the domain $\mathbb{N}$ of the Peano system $(\mathbb{N}, 0, +1)$ which, if we use second-order rather than first-order logic, as a consequence of Dedekind's recursion theorem, is isomorphic to any Peano system (N, o, S). Of course, $+1 : \mathbb{N} \to \mathbb{N}$ such that $+1(n) \mapsto n + 1$, for any $n \in \mathbb{N}$. See on this Drake and Singh (1996), Chap. 6, Sect. 6.2.11.

[36] In first-order ZFC we can only deal with small structures.

[37] A class K is transitive if $x \in \mathrm{K}$ implies $x \subset \mathrm{K}$.

of this fact are very important, because if first-order ZFC is a formal system of a theory of structures then the nature of the elements of any domain M of a model (M, ∈) of first-order ZFC would be mathematically irrelevant. But, of course, if this is the case we have that the project of a Russellian foundation of mathematics, that is, the project of individuating the prime/basic sets that first-order ZFC is intended to describe and from which it is possible to reconstruct any object studied by any mathematical theory T reducible to/expressible within first-order ZFC is doomed to failure. (The SPF* of Sect. 9.2 cannot be solved.)

Now, given that in Sect. 9.2 we answered the First Problem of Foundations—Is it possible to establish once and for all the certainty of the methods adopted in mathematics?—in the negative, and that in this section we also answered in the negative the Second Problem of Foundations relativized to first-order ZFC (SPF*), we need to ask what we should learn from these results.

The first thing we should learn is that the traditional foundational ideology that sees mathematics as an edifice resting on unshakeable foundations, an edifice whose stability, apart from the solidity of the foundations, is also guaranteed by the certainty of the methods adopted in its construction, is unwarranted.

Secondly, given that first-order ZFC is unable to: (a) establish once and for all the certainty of the methods adopted in mathematics; and (b) individuate the prime/basic sets from which it is possible to reconstruct any object studied by any mathematical theory T reducible to/expressible within first-order ZFC, then, perhaps, the best way of describing first-order ZFC is as an axiom system which, like the axiom systems for groups or topological spaces, is satisfied by a plurality of structures. This is a very important point, because if it is correct to say that first-order ZFC is like the axiom systems for groups and topological spaces then, as in the case of the axiom systems for groups and topological spaces there is no 'intended group' or 'intended topological space', in the same way in the case of first-order ZFC there is no intended universe or 'true V' either.

Thirdly, note that, in a context in which first-order ZFC 'has gone structural,' the belief in the existence of a multiplicity of non-isomorphic *universal structures* (models of first-order ZFC), is neither a vindication of the hyper-universe view we decided to put aside in Sect. 9.3 nor is a form of modal structuralism *à la* Hellman.

It is not a vindication of the hyper-universe view, because the form of structuralism defended in this paper rejects the very concept of universe as the collection of all sets. And it is not a form of modal structuralism *à la* Hellman either,[38] because, in contrast with Hellman's view on these matters, the type of structuralist view of mathematics here advocated is characterized by what we may call 'metaphysical realism' (see footnote 1, p. xxx).

Incidentally, in spite of the great interest of an attempt aimed at reconciling metaphysical anti-realism with realism about truth in mathematics (see footnote 1, p. xxx)

[38]For Hellman, '[M]athematics is the free exploration of structural possibilities, pursued by (more or less) rigorous deductive means' (Hellman 1989, Introduction, p. 6), where the notion of logical possibility 'functions as a primitive notion, and must not be thought of as requiring a set-theoretical semantics for it to be intelligible' (Hellman 1989, Chap. 2, p. 60).

and of the technical sophistication displayed in carrying out this project, Hellman's version of modal structuralism appears to be untenable.

In fact, since Hellman is not a metaphysical realist about structures, he stands in need of a justification for talking about possible structures. But, a justification for talking about possible structures can only come from (1) believing in an ontology of possible structures or from (2) producing a consistency proof of axiom systems like ZFC which are supposed to produce information about possible structures, and it is clear that Hellman can help himself neither to (1) nor to (2). He cannot use (1), because doing so would contradict the metaphysical anti-realism about structures typical of his version of modal structuralism; and, cannot take justificatory route (2) either as a consequence of Gödel's Second Incompleteness Theorem.

Fourthly, if by 'foundational research' we mean either what Hilbert called 'axiomatic thinking'[39] or, alternatively, the study of universal structures then the failure on the part of first-order ZFC to produce a positive answer to both the First and the Second Problem of Foundations cannot, obviously, have negative consequences on foundational research.

Fifthly, although realism about objects/sets is bankrupt as metaphysical component of the framework of a good foundational system like first-order ZFC, this does not mean that we should become anti-realists. For, such a traditional and inadequate version of realism (realism about objects/sets) can be replaced by another, and perfectly defensible form known in the literature as structural realism.

Sixthly, the unavoidable presence of rhetoric in mathematics (see Sect. 9.3), and the impossibility of giving a positive solution to the First Problem of Foundations within first-order ZFC present us with a counterintuitive view of mathematics in which, even though the a priori nature of mathematical justification—proof—stands unaffected, fallibility creeps in.

Lastly, the consequences of the dilemma which, as we have seen in Sect. 9.4, turn out to be unacceptable when this is formulated within a framework in which mathematics is a science of objects/sets, are easily accommodated within a structuralist representation of first-order ZFC. In fact, if, in our structuralist reading of first-order ZFC, this formal system behaves very much like the group axioms then, as in group theory there is nothing wrong with studying both abelian and non-abelian groups, there would be nothing wrong either in making the structures described by first-order

[39] According to Hilbert,

> [G]iven a mathematical theory T, it is possible to distinguish between results obtained within it which are of central importance—like the Fundamental Theorem of Calculus, the Fundamental Theorem of Algebra, the well-ordering theorem in set theory, the Completeness Theorem for first-order logic—and those which are not so important.
>
> Relatively to these fundamental results, the theory can be developed either *downwards* or *upwards*. Developing the theory downwards means deriving consequences from the fundamental results. Developing the theory upwards means finding some statements of the theory from which it is possible to derive the fundamental results. Axiomatic thinking is, for Hilbert, the way of regarding a mathematical theory from the point of view of what I going to call 'upward continuation'. [Oliveri (2005), p. 119.]

ZFC + CH and those described by first-order ZFC + ¬ CH an object of mathematical investigation of ‘first-order ZFC set theory’.

Furthermore, although, within a framework in which mathematics is a science of objects/sets, fist-order ZFC cannot provide a Russellian foundation of mathematics, if we consider first-order ZFC as a formal system of a theory which studies the properties of certain universal structures then first-order ZFC would regain a foundational rôle. For, in showing that the structuralist approach to mathematics can be smoothly and harmoniously carried out from groups, topological spaces, Peano systems, etc. all the way down to universal structures, first-order ZFC would, after all, provide some kind of Russellian justification to the idea that mathematics is a science of structures.

One thing we should notice in closing is that, since the traditional foundationalist ideology based on the architectonic metaphor is debunked and the best thing we can say about mathematics is that this is a science of structures then, perhaps, there is some room for an irenic solution of the long standing controversy between set theory and category theory over which of these two theories provides the correct foundations of mathematics. This is an irenic solution of the controversy, because since both set theory and category theory study structures of deep mathematical interest and importance, in a general perspective in which mathematics is seen as a science of structures, they are mutually compatible, and a new, most interesting problem comes to the fore, that of investigating the relationship existing between the structures they study.

9.6 Conclusions

In this paper we defended the centrality to both the philosophy of mathematics and mathematics of the question about whether or not there is such a thing as ‘the’ universe of sets, that is, the true V.

We argued that, if such a question is posed within a framework characterized by the idea that mathematics is a science of objects, it is possible to produce a dilemma each horn of which turns out to be unacceptable.

The way out of the dilemma we thought of taking consisted in the rejection of the idea that mathematics is a science of objects in favour of the view that mathematics is a science of structures.

The kind of structuralism about mathematics advocated in this paper, among other things, finds unacceptable the very idea of a universe as the collection/totality of all sets and, therefore, is against the idea of a true V; upholds metaphysical realism about structures; does not do away with (mathematical) objects, but restricts the significance of mathematical investigation to the study of structures; is committed to realism about truth in the sense that the truth/falsity of a mathematical statement S *in a given structure* is independent of the possibility of knowing which is the case; hints at a structuralist foundation of mathematics which is no longer dependent on what we have repeatedly called the ‘architectonic metaphor’.

Of course, in writing the article it would have been useful to have engaged in an extensive comparative study of the version of structuralism in mathematics here offered with old and new versions of mathematical structuralism which, including Hellman's, go from Bourbaki (1996), Shapiro (1997), and Resnik (2001) to the more recent views defended by the contributors to The Univalent Foundations Program (2013). But, as the reader will easily understand, this is not something that can be done in the short space offered by a paper.

References

Balaguer, M. (1998). Non-uniqueness as a non-problem. *Philosophia Mathematica*, *6*, 63–84.

Benacerraf, P. (1985). What numbers could not be. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics. Selected readings* (pp. 272–294). Cambridge: Cambridge University Press.

Bourbaki, N. (1996). The architecture of mathematics. In W. B. Ewald (Ed.), *From Kant to Hilbert* (pp. 1265–1276). Oxford: Clarendon Press.

Burali-Forti, C. (1967a). A question on transfinite numbers. In J. van Heijenoort (Ed.), *From Frege to Gödel* (pp. 104–111). Cambridge, Massachusetts: Harvard University Press.

Burali-Forti, C. (1967b). On well-ordered classes. In J. van Heijenoort (Ed.), *From Frege to Gödel* (pp. 111–112). Cambridge, Massachusetts: Harvard University Press.

Cantor, G. (1962). Cantor an Eulenburg. In E. Zermelo (Ed.), *Gesammelte Abhandlungen* (pp. 400–407). Hildesheim: Georg Olms Verlagsbuchhandlung.

Cellucci, C. (2000). *Le ragioni della logica*. Roma, Bari: Laterza.

Cellucci, C. (2005). Mathematical discourse vs. mathematical intuition. In C. Cellucci & D. Gillies (Eds.), *Mathematical reasoning and heuristics* (pp. 137–165). London: King's College Publications.

Drake, F. R., & Singh, D. (1996). *Intermediate set theory*. Chichester: Wiley.

Dummett, M. A. E. (1975). The philosophical basis of intuitionistic logic. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics. Selected readings* (pp. 97–129). Cambridge: Cambridge University Press.

Dummett, M. A. E. (1991). *The logical basis of metaphysics*. London: Duckworth.

Ferreirós, J., & Jeremy, G. (Eds.). (2006). *The architecture of modern mathematics*. Oxford: Oxford University Press.

Forti, M., Honsell, F., & Lenisa, M. (1996). Operations, collections and sets within a general axiomatic framework. *Quaderni del Dipartimento di Matematica Applicata "U. Dini"*, *32*, 1–22.

Frege, G. (1980). *The foundations of arithmetic* (Austin, J. L. Trans., 2nd revised ed.). Oxford: B. Blackwell.

Grosholz, E. (2007). *Representation and productive ambiguity in mathematics and the sciences*. Oxford: Oxford University Press.

Hamkins, J. D. (2011). *The set-theoretic multiverse*. arXiv:1108.4223v1. Aug 22, 2011.

Hellman, G. (1989). *Mathematics without numbers. Towards a modal-structural interpretation*. Oxford: Clarendon Press.

Hilbert, D. (1926). On the infinite. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics. Selected readings* (pp. 183–201). Cambridge: Cambridge University Press.

Isaacson, D. (1987). Arithmetical truth and hidden higher-order concepts. In W. D. Hart (Ed.), *The philosophy of mathematics* (pp. 203–224). Oxford: Oxford University Press.

Jech, T. (2003). *Set theory. The third millennium edition, revised and expanded*. Berlin, Heidelberg: Springer.

McLarty, C. (1995). *Elementary categories, elementary toposes*. Oxford: Clarendon Press.

Montague, R. M. (1955). On the paradox of grounded classes. *Journal of Symbolic Logic*, *20*(2), 140. http://projecteuclid.org/euclid.jsl/1183732123
Oliveri, G. (2005). Do we really need axioms in mathematics? In C. Cellucci & D. Gillies (Eds.), *Mathematical reasoning and heuristics* (pp. 119–135). London: King's College Publications.
Oliveri, G. (2006). Mathematics as a quasi-empirical science. *Foundations of Science*, *11*, 41–79.
Oliveri, G. (2007). *A realist philosophy of mathematics*. London: College Publications.
Oliveri, G. (2011). Productive ambiguity in mathematics. *Logic & Philosophy of Science*, *9*, 159–164.
Oliveri, G. (2012). Object, structure, and form. *Logique & Analyse*, *219*, 401–442.
Peano, G. (1889). Arithmetices principia nova methodo exposita. In J. van Heijenoort (Ed.), *From Frege to Gödel* (pp. 83–97). Cambridge, Massachusetts: Harvard University Press.
Rav, Y. (2008). The axiomatic method in theory and in practice. *Logique & Analyse*, *202*, 125–148.
Resnik, M. D. (2001). *Mathematics as a science of patterns*. Oxford: Clarendon Press.
Shapiro, S. (1997). *Philosophy of mathematics. Structure and ontology*. Oxford: Oxford University Press.
Tuller, A. (1967). *A modern introduction to geometries*. New York: Van Nostrand.
The Univalent Foundations Program. (2013). *Homotopy type theory. Univalent foundations of mathematics*. Princeton: Institute for Advanced Study. https://hottheory.files.wordpress.com/2013/03/hott-a4-611-ga1a258c.pdf
Wittgenstein, L. (1983). *Philosophical investigations*. Oxford: B. Blackwell.

Chapter 10
The Search for New Axioms in the Hyperuniverse Programme

Sy-David Friedman and Claudio Ternullo

Abstract The Hyperuniverse Programme, introduced in Arrigoni and Friedman (2013), fosters the search for new set-theoretic axioms. In this paper, we present the procedure envisaged by the programme to find new axioms and the conceptual framework behind it. The procedure comes in several steps. Intrinsically motivated axioms are those statements which are suggested by the standard concept of set, i.e. the 'maximal iterative concept', and the programme identifies higher-order statements motivated by the maximal iterative concept. The satisfaction of these statements ($\mathbb{H}$-axioms) in countable transitive models, the collection of which constitutes the 'hyperuniverse' ($\mathbb{H}$), has remarkable first-order consequences, some of which we review in Sect. 10.5.

Keywords New axioms · Set-theoretic multiverse · Hyperuniverse programme

10.1 New Set-Theoretic Axioms

Over the last years, there has been an intense debate within the set-theoretic community concerning the acceptance or non-acceptance of several set-theoretic statements such as V= L, large cardinals, axioms of determinacy (AD, PD, $AD^{L(\mathbb{R})}$) or forcing axioms (MA, PFA, etc.) and the discussion seems to be nowhere near being settled.

Sy-D. Friedman (✉) · C. Ternullo
Kurt Gödel Research Center for Mathematical Logic, Vienna, Austria
e-mail: sdf@logic.univie.ac.at

C. Ternullo
e-mail: claudio.ternullo@univie.ac.at

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_10

The received view concerning an axiom is that it should be ‘self-evident’, i.e., that it should be immediately, and with little effort, acknowledged as true. If such a view is still to be held, then there is no hope to accept the aforementioned statements as new axioms.[1]

But even if one discards the ‘self-evidence view’ as inapplicable, there are still deep issues which have to be addressed by anyone supporting the acceptance of one or more of the statements mentioned above and, more generally, of any axiom candidate.

First of all, there is often a lack of intrinsic motivation for such statements, where, by ‘intrinsic’, as explained at length in the sections below, we mean ‘required by the concept of set’. Secondly, the view that a new axiom should be accepted as true of the realm of sets has been seriously challenged by the independence phenomenon and the related existence of a set-theoretic multiverse: it is often relatively easy to produce a universe of sets which contradicts a given set-theoretic statement. Finally, all new axiom candidates are first-order and one main worry we want to bring out in this paper is precisely that first-order principles may be too weak to capture further properties of the cumulative set-theoretic hierarchy.

One clear preliminary upshot of the informal considerations above is the following: it is unlikely that any new first-order axiom candidate will be accepted on its own as an intrinsically motivated principle of set theory. Granted, it might still be accepted on purely extrinsic grounds, but it is not clear that this would be sufficient evidence for its acceptance.

In this paper, we are going to propose an alternative way to identify new intrinsically motivated set-theoretic axioms, which originates from the conceptual framework of the Hyperuniverse Programme, as detailed in Arrigoni and Friedman (2013), and which fosters a revisionary conception of what a ‘new’ axiom is. In our view, new axioms are higher-order set-theoretic principles, more specifically principles expressing the maximality of the universe of sets (V). The latter are strong mathematical propositions, some of which have been gradually isolated and examined in recent years in work by the first author and others, and, more recently, by the first author and Honzik.[2] We believe that there is a sense in which such propositions, as will be presented in Sect. 10.5, can legitimately claim to be motivated by the concept of set and, by virtue of this, be viewed as intrinsically motivated new axioms.

It should be mentioned that all of these statements have striking first-order set-theoretic consequences, which we will describe in more detail in the next sections

[1]Of course, it is also as much debatable that the standard axioms of set theory, that is, ZFC, are all ‘self-evident’. A very natural case in point is the Axiom of Choice, but one may have equally reasonable reservations on the Axioms of Infinity, Replacement or Foundation. A thorough discussion of some of these issues can be found in Wang (1974), Maddy (1988a, b) and Potter (2004).

[2]See, in particular, Friedman (2006), Arrigoni and Friedman (2012, 2013) and Friedman and Honzik (2016).

and this fact, although not representing an intrinsic justification for their acceptance, indisputably adds to their mathematical attractiveness.

One further goal of the Hyperuniverse Programme is to find one single 'optimal' maximality principle, whose acceptance would, thus, lead to identifying one single collection of first-order consequences. Therefore, our foundational project fosters the view that the procedure described here might also count as a procedure to find solutions to the open problems of set theory. However, the notion of 'solution', here, is inevitably as much revisionary as that of 'new' axiom.

The structure of the paper is as follows. In Sects. 10.2 and 10.3, we briefly discuss the features of 'intrinsic' evidence and set forth our conception of the set-theoretic universe as being a 'vertical' multiverse. In Sect. 10.4, we introduce the hyperuniverse as our auxiliary multiverse, wherein one can investigate the consequences of the maximality of V, through the use of V-logic. In Sect. 10.5, we enunciate maximality principles for V which, in our view, are motivated by the concept of set. In Sect. 10.6 we discuss the notion of 'new' axiom *qua* $\mathbb{H}$-axiom, then, in Sect. 10.7, we proceed to make some final considerations. Finally, in Appendix, we present some results which show that a heavily investigated collection of new set-theoretic axioms, absoluteness axioms, which has recently received a lot of extrinsic support, may fall short of the requirements described above.

10.2 Intrinsic Evidence for New Axioms

10.2.1 Brief Remarks on Ontology and Truth

In the next pages, we will be making frequent reference to issues of ontology and truth and it is maybe appropriate to briefly address these issues before examining the notion of intrinsic evidence.

On the grounds of what the programme aims to yield, i.e. new set-theoretic axioms, it is entirely natural to ask whether new axioms should be seen as 'true' statements of set theory, and in what sense. We will make it clear in the next section in what sense they should be viewed as 'true', but first we want to say something more general about 'truth'.

The programme's position is that axioms do not reflect truth in an independent realm of mathematical entities. It is rather the *concept of set* that plays a key role in our foundational project. As we shall see in the next subsection, and as is commonly acknowledged in set theory, the concept of set is instantiated by a specific mathematical structure, the *cumulative hierarchy*, but it does not automatically provide us with a fully determinate collection of properties of sets in this structure. Now, we believe that it is possible to derive properties of the concept of set which provide us with an indication of what further properties the set-theoretic hierarchy should have.

There is possibly a hint of realism in this position, insofar as we view the concept of set as being a 'stable' feature of our experience of sets and we subscribe to its stability in the sense that we do not question the ZFC axioms which are true of it.

However our view follows an overall epistemological concern, that of securing the truth of new axioms and of their first-order consequences through setting forth an alternative evidential framework for them which does not imply a pre-formed ontological picture. Therefore, ontology, in the most robust sense of the word, does not play a pre-eminent role in our project.

If there is a detectable ontological framework within our account, that is the *core* structure we identify in Sect. 10.3, i.e. the tower-like multiverse of V_κ's, where κ is a strongly inaccessible cardinal. In turn, properties of this multiverse will motivate the adoption of one further ontological construct, the hyperuniverse, which consists of all countable transitive models. Neither multiverse is given *a priori*.

The maximality principles we will be concerned with quantify over extensions of V. However, our language is that of first-order ZFC, therefore maximality principles do not formally involve talk of classes. So, in the end, we have sets, and nothing else.

This ontological view might be seen as entailing a conception of truth that lacks the requisite strength to see axioms as 'true'. But in fact, as we will see, the concept of set is adequate to make strong claims about set-theoretic maximality, for instance alternative conceptions of *vertical* maximality are ruled out as unwarranted on the grounds of the concept itself. It is true, however, that in order to have models where new axioms 'live', one has to shift to countable transitive models and, thus, to a different framework of truth. But this is not overall necessary. One can still appreciate the force of maximality principles within the whole V and, thus, stick to a vision of truth and ontology entirely befitting the concept of set. Further details on our positions will be given in the next few sections.

10.2.2 Two Sources of Evidence

Although the distinction is not entirely perspicuous, since Gödel (1947), it has become fairly commonplace in the literature to refer to two main different forms of evidence for the acceptance of an axiom as 'intrinsic' (internal) and 'extrinsic' (external) evidence. Very roughly, the distinction can be glossed as follows. Intrinsic evidence for an axiom is that following from the concept of set, whereas extrinsic evidence relates to the fruitfulness and success of an axiom, possibly also outside set theory. In other terms, an axiom may be accepted either because it expresses a 'necessary' property of sets or because it is corroborated by good results (and interesting practice) or for both reasons.

The issue of whether this distinction has any plausibility is beyond the scope of this article and, for the sake of our arguments, we will not challenge it. However, it should be noticed that, in our opinion, in opposition to the point of view expressed by some authors, 'intrinsicness' does not imply the view we have just informally rejected, that axioms should be 'self-evident'.[3] In fact, an axiom may be true of the concept of set and not be immediately graspable as true. This is because not all true properties of the concept of set are immediately graspable. Therefore, arguably, it is our task to gradually uncover such properties, by clarifying the content of our intuitions. Ideally, we should be able to determine the properties of the concept of set, and possibly of other properties of clear set-theoretic relevance, by following what Potter calls the 'intuitive' method:

> The intuitive method invites us instead to clarify our understanding of the concepts involved to such an extent as to determine (some of) the axioms they satisfy. The aim should be to reach sufficient clarity that we become confident in the truth of these axioms and hence, but only derivatively, in their consistency. If the intuitive method is successful, then, it holds out the prospect of giving us greater confidence in the truth of our theorems than the regressive method.[4]

But just what form of *intuition* does the intuitive method presuppose? Is intuition alone sufficient to justify the adoption of set-theoretic axioms? These are no doubt vexing questions on which we cannot fully dwell in this paper. However, some considerations are in order.

Intuition is, sometimes, construed in the Gödelian sense, as a faculty of perception which provides us with detailed information on mathematical objects. However, as we have seen, the programme does not commit itself to any form of object-realism. Therefore, our appeal to intuition and the intuitive method should be construed in the following way: as hinted at by Potter, we seem to have the ability to single out the relevant concepts and properties that are derivable from the concept of set.

As we shall see, the cumulative hierarchy instantiates the concept of set (as described below) and its maximality seems to follow naturally. Now, does that mean that we need to have access to platonistic entities in order to successfully carry out this task? We do not have a definite answer to this question, but, on the grounds of the considerations made in the previous section, it seems natural to tentatively rule out such possibility: realism should not extend so far as to postulate the existence of an independent, pre-formed ontology, but rather only postulate a stable concept of set, from which further properties of sets can be derived.

It should be noticed that we do not hold that our maximality principles, such as the IMH, become thus straightforwardly 'intrinsically justified'. What we believe to be intrinsically justified by the concept of set is rather the feature of the maximality of the cumulative hierarchy and, consequently, its *maximal extendibility*. Maximality principles take different forms, so we could, at most, say that such forms are *intrinsi-*

---

[3] For instance, the equating of 'intrinsicness' with 'self-evidence' is clearly hinted at in the following passage of Maddy (1988a), p. 482: 'The suggestion is that the axioms of ZFC follow directly from the concept of set, that they are somehow 'intrinsic' to it (obvious, self-evident) [...]'.

[4] The 'regressive', as opposed to 'intuitive', method mentioned by Potter holds that '...the object of a good axiomatization is to retain as many as possible of the naive set-theoretic arguments which we remember with nostalgia from our days in Cantor's paradise, but to stop just short of permitting those arguments which lead to paradox' (Potter 2004, p. 36).

cally motivated, insofar as maximality, in general, is an intrinsically justified feature of the concept of set.

There is, of course, further work to be done to establish the stronger claim that some of our maximality principles are intrinsically justified, and we can only hope that further intrinsic evidence may, one day, substantially help bolster this claim.

10.2.3 The Maximum Iterative Concept

Although it is not clear what Cantor took 'sets' to be at the beginning of his set-theoretic investigations, over the years increasingly wide agreement has been reached that the concept implies an account of the iterative formation of all sets along stages indexed by the ordinals. By this account, each set belongs to a stage of what has been called the *cumulative hierarchy*, starting with the empty set, and then iterating the power-set at all successor stages and the union of all sets formed in previous stages at limit stages of the hierarchy.

In fact, the iterative conception is more correctly referred to as the 'maximal iterative conception':

[MIC] (1) All sets which can be formed at each stage are actually formed. (2) The formation of sets should continue as far as possible.

The vocabulary used in the formulation of the [MIC] has been variously interpreted as hiding *temporal*, *modal* and, in general, *metaphysical* forms of mutual dependency among sets, elements and stages, and this aspect is responsible for some sort of conceptual opacity in the [MIC].[5]

However, leaving aside such troubles for the time being, it seems clear that the basic rationale underlying the [MIC] is that the procedures to form sets ought not to be constrained by 'internal' limitations, that is, by mathematical principles hindering the maximisation and the continuation of such formation. This line of thought has been distinctly referred to in a fortunate article by Bernays as 'quasi-combinatorialism', the conceptual attitude which would allow one to treat and manipulate all mathematical objects, both finite and infinite, and combinations thereof, as fully determinate objects of thought.[6] The maximal character of the [MIC], therefore, can be motivated using a 'quasi-combinatorial' conceptual framework, as, by this, one does not put any constraint on the class of producible sets.

These are well-known facts. Now, we want to take a step further. The intuitive method invites us to focus our attention on one specific feature of the [MIC] that we are going to use extensively in the rest of the paper. Suppose one takes the cumulative

[5] For an exhaustive overview of these issues see, again, Potter (2004), in particular, pp. 34–41, or Jané (2005b).

[6] In Bernays' own words, 'quasi-combinatorialism' ultimately refers to '...an analogy of the infinite with the finite.' See Bernays (1935), reprinted in Benacerraf and Putnam (1983), p. 259.

hierarchy to be a determinate object of thought, V, the universe of sets. Then the [MIC] may also imply one further principle of 'plenitude', which can be formulated in the following way:

[MaxExt] Given a universe of sets, all possible extensions of it which can be formed are actually formed.

The 'extensions' referred to in [MaxExt] are given by the creation of 'new' sets in the only two ways we know, either adding new stages to the hierarchy or 'producing' new subsets at successor stages; thus the principle seems to be perfectly justified in light of the [MIC]. However, the principle seems to shed light on one further dimension in the iteration, insofar as it assumes that the latter should get us beyond the universe itself. This automatically introduces the issue of whether we have grounds to believe that the universe is a determinate (actual) object of thought. As we shall see, there is, in fact, a way to interpret [MaxExt] in a way which remains faithful to its nature, but does not imply this kind of actualism.

In any case, our goal, for the time being, is to re-state the notion of 'intrinsicness' in the following way: intrinsic evidence for the acceptance of an axiom is that related to the [MIC], which, in particular, implies [MaxExt] as one of its features.

10.3 Conceptions of V. The 'Vertical' Multiverse

10.3.1 What Is V? The Actualism/Potentialism Dichotomy

[MaxExt] seems to imply that the universe can be 'extended' and that there is no limitation on how much it can be extended. Extensions of the universe, as we said, are given by further stages in the cumulative hierarchy or new subsets. However, as anticipated, there is a difficulty in this point of view: the literal sense of the notion of 'extension', implies, at the very least, that what is extended is an object with 'boundaries', that is, a 'delimited' object. Now, it is not clear that the cumulative hierarchy is one such object. As a matter of fact, in the [MIC] there is nothing which commits us to seeing V as a delimited object. On the contrary, it would seem that V is best construed as an open-ended sequence of stages. On the other hand, the standard interpretation of the first-order quantifiers is that they range over the class of *all* sets, as though *all* sets were made available to us by unbounded quantification. So, what (if any) is the fact of the matter?

Debates over the nature of the infinite, whether it be *actual* or *potential* or both, date already to antiquity. The two viewpoints we have summarised above extend this kind of debate to the nature of one specific instance of the infinite, the universe of all sets. *Potentialists* believe that this object is neither actual nor actualisable, whereas *actualists* do.[7]

Now, [MaxExt] seems to fit best the potentialist viewpoint, insofar as it posits the existence of extensions of V. For, if V is an actualised domain, how could it possibly be extended?

Therefore, on the potentialist viewpoint as befitting [MaxExt], one cannot even speak of V, as there is no such actualised and determinate object as V, but one should rather refer to an endless sequence of initial segments V_α's whose union can always be extended. This means that [MaxExt] implies, at least, that the cumulative hierarchy is open-ended and that new stages in the formation of sets can always be formed. However, as we have seen, this is just one way of extending the universe. Extensions of V are not only extensions of its *height*, but also of its *width*. The width of the universe is given by the power-set operation and, thus, an extension of V in width means that there is also a way to expand the range of the power-set operation.

So, now we have a more complex picture. One could be either an actualist or a potentialist in either height or width, as summarised below:

HEIGHT ACTUALISM: the height of V is fixed, that is, no new ordinals can be added,
WIDTH ACTUALISM: the width of V is fixed, that is no new subsets can be added.

HEIGHT POTENTIALISM: the height of V is not fixed, new ordinals can always be added,
WIDTH POTENTIALISM: the width of V is not fixed, new subsets can always be added.

and combinations thereof.

As said, [MaxExt] seems to commit us to a full-blown form of potentialism, both in height and width. This makes full sense, especially from the point of view of 'quasi-combinatorialism': by this attitude, no internal limitation of the procedures to form new sets should be applicable and this also extends to such large-scale objects as V.

However, there is one difficulty with this view. While height potentialism seems to be robustly supported by our idea of 'adding' new ordinal-indexed stages to the cumulative hierarchy, so that we can always form a sequence of V_α's increasing with α, it is far more problematic to see how extensions of the width of the universe may come in 'stages'. In fact, such extensions as, for instance, the possible set-generic extensions of the universe are not organised in stages at all.

[7] An examination of the potentialist and actualist positions, with reference to the justification of *reflection principles*, is carried out in Koellner (2009), which also draws upon Tait (1998a, b). A thoroughly actualist point of view on reflection is expounded in Horsten and Welch (2013). A potentialist conception is described in Linnebo (2013), which provides a modal account of the axioms of set theory already explored in Hellman (1989) and Parsons (1983). For the early debate on such issues as the nature of the universe of sets, the role of the absolute infinite and proper classes, all of which are relevant to the *actualism/potentialism* debate, also see the indispensable Hallett (1984), as well as Jané (1995), and Wang (1996), which contains Gödel's late conceptions on V.

Therefore, whereas our intuitions about the [MIC] seem to suggest that the universe is a fully potential hierarchy of sets, in both height and width, it could be argued that it is simply not possible to make sense of extensions of the universe in width in a way which is in line with the iterative, stage-like character of the [MIC].

The Hyperuniverse Programme has recently fostered a conception which acknowledges the significance of this objection,[8] and that, therefore, follows a conception alternative to full-blown potentialism which historically was first brought forward by Zermelo. We now proceed to briefly review Zermelo's conception.

10.3.2 Zermelo's Account: A 'Vertical' Multiverse

As is known, in his seminal paper Zermelo (1930), Zermelo investigates 'natural models' of his axioms, that is, models indexed by *boundary numbers* (fixed ordinals). Zermelo also proves that natural models form a linear hierarchy by inclusion. An example of a natural model of ZFC is given by V_κ, where κ is a *strongly inaccessible* cardinal.

Now, as said, we construe Zermelo's position as a specific one in the actualism *vs* potentialism debate: the Zermelian account is *potentialist in height* and *actualist in width*.

Zermelo's actualism in width follows from the presence of second-order quantifiers in (some of) his axioms. In fact, Zermelo's 1930 axiomatisation is, essentially, second-order. It is this fact that allows him to establish the quasi-categoricity of set theory or, in more rigorous terms, that:

Theorem 1 *Given any two extensional and well-founded structures M_1 and M_2, such that $M_1 \models Z_2$ and $M_2 \models Z_2$ (where Z_2 denotes the axioms of second-order set theory), only three cases can occur: M_1 is isomorphic to M_2, M_1 is isomorphic to a proper initial segment of M_2, or M_2 is isomorphic to a proper initial segment of M_1.*

A trivial consequence of quasi-categoricity is the absoluteness of the power-set operation, which automatically leads one to width actualism. However, our emphasis, here, is on the 'quasi-' bit of his result, since models may still differ in height and, thus, be extendible in a way which clearly suggests height potentialism. In particular, Zermelo construed the sequence of V_α's as stopping points in an *endless* process of potentialisation of an only temporarily actualised universe.

Zermelo vividly recapitulates his approach in the following manner:

> To the unbounded series of Cantor's ordinals there corresponds a similarly unbounded double-series of essentially different set-theoretic models, in each of which the whole classical theory is expressed. The two polar opposite tendencies of the thinking spirit, the idea of creative advance and that of collection and completion [*Abschluss*], ideas which also lie

[8] See, in particular, Antos et al. (2015) and Friedman (n.d.).

> behind the Kantian 'antinomies', find their symbolic representation and their symbolic reconciliation in the transfinite number series based on the concept of well-ordering. This series reaches no true completion in its unrestricted advance, but possesses only relative stopping-points, just those 'boundary numbers' [*Grenzzahlen*] which separate the higher model types from the lower. Thus the set-theoretic 'antinomies', when correctly understood, do not lead to a cramping and mutilation of mathematical science, but rather to an, as yet, unsurveyable unfolding and enriching of that science. (Zermelo 1930, in Ewald 1996, p. 1233)

Zermelo's sequence of natural models can also be viewed as a tower-like multiverse, a 'vertical' multiverse, a collection of universes linearly ordered by inclusion.

Unfortunately, at the practical level, the 'vertical' multiverse fits only half of [MaxExt]: extensions in height are now incorporated within this picture, whereas extensions in width are banned. However, as we said, this seems to be more in line with some worries concerning the impossibility, from a mathematical point of view, to account for extensions in width in an orderly fashion.

Therefore, if we want to keep full potentialism and Zermelo's account, we have to find a way to address also extensions in width within this account. This task we accomplish in the second half of the next section, by introducing V-logic.

10.4 The Hyperuniverse ($\mathbb{H}$). V-Logic

In the previous sections we have established two facts: (1) intrinsic evidence relates to the [MIC], in particular, to one of its features, that is [MaxExt]; (2) as we have seen, [MaxExt] seems to be more in line with a full-blown potentialist picture of the universe. However, there is no way to address extensions of the width of the universe in a way which suits the iterative character of the [MIC], therefore we ought to settle on an account of V wherein the width of the universe is fixed. Such an account is very fittingly provided by the Zermelian 'vertical' multiverse.

Before turning to the programme's maximality principles in the next section, we first have to carry out two tasks: we have to show that there is indeed a way to formulate principles addressing extensions of the universe not only in height but also in width within a Zermelian conceptual framework and, secondly, we have to identify universes where first-order consequences of such principles hold. We start with the latter goal: the hyperuniverse provides an ontological environment where one can investigate consequences of our maximality principles.

10.4.1 The Hyperuniverse

Let us leave aside, for a moment, the concept of set, the [MIC] which constitutes its full expression, the ensuing picture of the realm of sets as the cumulative hierarchy and let us turn our attention to the techniques used by set-theorists to establish results concerning set-theoretic truth.

As is known, there is only one way to establish the independence of set-theoretic statements from the axioms, i.e. through finding two models wherein that statement and its negation are, respectively, true. If the axioms are consistent, then they cannot prove or disprove such a statement.

There is a wide variety of models that set-theorists investigate: e.g., the constructible universe L, core models K, HOD, $M[G]$ (where G is a generic filter on a forcing poset $\mathbb{P} \in M$) and so forth. The main techniques employed consist in the construction of an *inner model* and of a *forcing extension* of a ground model M. Almost invariably, the ground model used is a *countable transitive model*.

So, the problem is the following: how do all these models relate to the concept of set, which seemed to give rise to a unique picture of the realm of sets, that is, V? Moreover, does each of these models constitute a separate and, to some extent, alternative ontological construct?

The situation we are to face up to here is direly ambivalent. On the one hand, one could legitimately claim that all model-theoretic constructions are in V, 'reflecting' the universe each in its own particular way. On the other hand, one could say that, if V is a fully determinate construct, something which seems plausible in light of our adoption of the [MIC] and of its associated Bernaysian 'quasi-combinatorialism', then all of these models represent different and, sometimes, mutually incompatible versions of set-theoretic truth, which cannot possibly be amalgamated into one single framework.

Now, call the view that there is a single universe of sets *monism*, whereas let *pluralism* be the view that there are many universes, and that V has no ontological priority. Our approach is alternative to both and may be legitimately called 'dualistic'. Within the programme, we are, in a sense, forced to postulate both the existence of one 'extendible' universe and, at the same time, that of a plural framework containing many universes, where properties of the universe allow the detection of further set-theoretic truth. Now, the models we want to confine our attention to are countable transitive models and our plural framework is defined as follows:

Definition 1 (*Hyperuniverse*) Let $\mathbb{H}^{ZFC}$ be the collection of all countable transitive models of ZFC. We call $\mathbb{H}^{ZFC}$ the hyperuniverse.[9]

But just why should one confine one's attention only to countable transitive models? Our choice is not related to the concept of set and rather originates from concerns arising from practice: we want to infer new truth (first-order statements) from intrinsically motivated new axioms (maximality principles) and, in order to do this, countable transitive models are not only suitable, but also necessary (more details on this are given below in our discussion of V-logic). Further reasons for adopting

[9]Henceforth, we shall only use $\mathbb{H}$ to refer to it.

the hyperuniverse as a multiverse construct are more precisely substantiated in what follows:

(1) First of all, it should be noticed that $\mathbb{H}$ is *closed under forcing and inner models*, which, as we saw, are the main techniques in the current practice. In other terms, if we start with countable transitive models, the use of forcing and inner models does not require more than and leave us with countable transitive models.
(2) The satisfaction of maximality principles in countable transitive models is also already suggested by the Löwenheim–Skolem theorem: given a statement ϕ, if ϕ is true in V, then ϕ is true in *some* element of the hyperuniverse. However, the notion of 'satisfaction', here, has to be mathematically secured more robustly (see Sect. 10.4.2 below).
(3) In $\mathbb{H}$, as a consequence of its very definition, there is no ill-founded model, and this fact is perfectly in line with our motivating evidential framework, that is, the [MIC].

Therefore, the adoption of the hyperuniverse is entirely subservient to achieving the result we wish to attain, that of finding new set-theoretic truth, but, as we have seen, is also well justified in light of different concurrent considerations and, in particular, of the fact that countable transitive models constitute the main tool used by set-theorists to investigate set-theoretic truth, a tool whereby the iterative and well-founded character of the cumulative hierarchy expressed by the [MIC] can be very aptly reproduced in a small-scale context.

10.4.2 V-Logic

We now proceed to describe how one can make sense of width maximality using V-logic. Such width maximality principles include the IMH, SIMH, IMH# and SIMH#, all of which will be defined in the next section.

As we said, the Löwenheim–Skolem theorem allows one to argue that any first-order property of V reflects to a countable transitive model. However, on a closer look, one needs to deal with the problem that not all relevant properties of V are first-order over V. In particular, the property of V 'having an outer model (a 'thickening') with some first-order property' is a higher-order property. We show now that, with a little care, all reasonable properties of V formulated with reference to outer models are actually first-order over a slight extension ('lengthening') of V.

We first have to introduce some basic notions regarding the infinitary logic $L_{\kappa,\omega}$, where κ is a regular cardinal.[10] For our purposes, the language is composed of κ-many variables, up to κ-many constants, symbols $\{=, \in\}$, and auxiliary symbols. Formulas in $L_{\kappa,\omega}$ are defined by induction: (i) All first-order formulas are in $L_{\kappa,\omega}$;

[10]Full mathematical details are in Barwise (1975). We wish to stress that the infinitary logic discussed in this section appears only at the level of theory as a tool for discussing outer models. The ambient axioms of ZFC are still formulated in the usual first-order language.

(ii) Whenever $\{\varphi\}_{i<\mu}$, $\mu < \kappa$ is a system of formulas in $L_{\kappa,\omega}$ such that there are only finitely many free variables in these formulas taken together, then the infinite conjunction $\bigwedge_{i<\mu} \varphi_i$ and the infinite disjunction $\bigvee_{i<\mu} \varphi_i$ are formulas in $L_{\kappa,\omega}$; (iii) if φ is in $L_{\kappa,\omega}$, then its negation and its universal closure are in $L_{\kappa,\omega}$. Barwise developed the notion of proof for $L_{\kappa,\omega}$, and showed that this syntax is complete, when $\kappa = \omega_1$, with respect to the semantics (see discussion below and Theorem 2).

Let us now consider a special case of $L_{\kappa,\omega}$, the so-called *V-logic*. Suppose V is a transitive set of size κ. Consider the logic $L_{\kappa^+,\omega}$, augmented by κ-many constants $\{\bar{a}_i\}_{i<\kappa}$ for all the elements a_i in V. In this logic, one can write a single infinitary sentence which ensures that if M is a model of this sentence (which is set up to ensure some desirable property of M), then M is an outer model of V (satisfying that desirable property). Now, the crucial point is the following: if V is countable, and this sentence is consistent in the sense of Barwise, then such an M really exists in the ambient universe.[11] However, if V is uncountable, the model itself may not exist in the ambient universe, but, in that case, we still have the option of staying with the syntactical notion of a consistent sentence.[12]

We have to introduce one further ingredient, that of an *admissible set*. M is an admissible set if it models some very weak fragment of ZFC, called Kripke–Platek set theory, KP. What is important for us here is that for any set N, there is a smallest admissible set M which contains N as an element—M is of the form $L_\alpha(N)$ for the least α such that M satisfies KP. We denote this M as $\mathrm{Hyp}(N)$.

And we have the following crucial result:

Theorem 2 (Barwise) *Let V be a transitive set model of ZFC. Let $T \in V$ be a first-order theory extending ZFC. Then there is an infinitary sentence $\varphi_{T,V}$ in V-logic such that following are equivalent:*

(1) *$\varphi_{T,V}$ is consistent.*
(2) $\mathrm{Hyp}(V) \models$ *"$\varphi_{T,V}$ is consistent."*
(3) *If V is countable, then there is an outer model M of V which satisfies T.*

By Theorem 2, if we wish to talk about outer models of V ('thickenings', that is, extensions of the width of V), we can do it in $\mathrm{Hyp}(V)$—a slight lengthening of V—by means of theories, without really thickening our V, that is, without postulating that such extensions are real. However, if we wish to have models of the resulting consistent theories, then, using the Löwenheim–Skolem theorem, we can shift to countable transitive models. And this is precisely where the hyperuniverse comes into play.

Now, we also want to make sure that members of the hyperuniverse really witness statements expressing the width maximality of V. One such statement is the Inner Model Hypothesis or IMH (for whose full examination see next section).

[11] Again, for more details we refer the reader to Barwise (1975).

[12] This means that the hyperuniverse, although fully justifiable in view of the use of V-logic, can be disposed of, if one only wants to keep the Zermelian multiverse (and its immediate connection with the [MIC] and [MaxExt]).

V satisfies the IMH if for every first-order sentence ψ, if ψ is satisfied in some outer model W of V, then there is a definable inner model $V' \subseteq V$ satisfying ψ. The formulation of IMH requires the reference to all outer models of V, but with the use of infinitary logic, we can formulate IMH syntactically in $\mathrm{Hyp}(V)$ as follows: V satisfies IMH if for every $T = ZFC + \psi$, if $\varphi_{T,V}$ from Theorem 2 above is consistent in $\mathrm{Hyp}(V)$, then there is an inner model of V which satisfies T. Finally, with an application of the Löwenheim–Skolem theorem to $\mathrm{Hyp}(V)$, this becomes a statement about elements of the hyperuniverse.

10.5 Maximality Principles for *V*

We have now arrived at the crux of our paper. Within the programme, we cast our new axioms as maximality principles about V and, after having established, using the notion of satisfaction in V-logic, that (1) these principles can be formulated in a Zermelian framework and (2) they are satisfied by members of $\mathbb{H}$, we can also see what first-order consequences they have through the study of countable transitive models, i.e. elements of the hyperuniverse.

First, there is one point which should be emphasised again: as the reader will see in a moment, the maximality principles that have been formulated within the programme all address extensions of V and, therefore, in our view, they specify ways such extensions, as postulated by [MaxExt], should be conceived of. Thus all such principles can be seen as specifications of [MaxExt]. As our evidential framework for the search for new axioms was given by the [MIC] and these principles follow from this evidential framework quite naturally, we believe that we have in this way found a source for new axioms based on the maximal interative conception.

Predictably, some principles refer to extensions in height and others to extensions in width. Accordingly, we may say that the former address the vertical maximality and the latter the horizontal maximality of the universe.

Vertical maximality has been recently formulated by the first author and Honzik in terms of a strong form of reflection called #-*generation*. We do not discuss the details here, but refer the reader to their paper Friedman and Honzik (2016).

Let us instead examine horizontal maximality. In the programme, this property is expressed by the IMH.

Definition 2 (*IMH*) If for every first-order sentence ψ, if ψ is satisfied in some outer model W of V, then there is a definable inner model $V' \subseteq V$ satisfying ψ.

Just to make things as clear as possible, 'outer models', in the definition above, are precisely the formal equivalent of extensions of the universe in width. Moreover, in our view, IMH prescribes the maximality of the universe (by using the language of 'extensions'), insofar as it prescribes its maximality with respect to *inner models*. Universes satisfying the IMH exist in $\mathbb{H}$:

Theorem 3 *Assuming the consistency of large cardinals, there are members of the hyperuniverse which satisfy the IMH.*

The proof is in Friedman et al. (2008), where it is shown that the consistency of slightly more than the existence of a Woodin cardinal is sufficient. One might question the use of Woodin cardinals here, which may not be intrinsically justified. But note that it is not the *existence* of Woodin cardinals that is needed to obtain the existence of members of $\mathbb{H}$ satisfying the IMH. It is only the consistency of Woodin cardinals that is used as an *auxiliary mathematical tool* in order to construct universes satisfying IMH and we believe that this fact does not commit us to asserting the existence of such cardinals, as 'consistency' is far less than 'existence'. It should be noted, incidentally, that in all members of $\mathbb{H}$ satisfying IMH there are *no* large cardinals at all. Therefore, if one believes that IMH is a correct higher-order principle about V, then one obtains that there are no large cardinals in V.
But the IMH does not take vertical maximality into account. Let IMH# denote the IMH for vertically-maximal, i.e. for #-generated, universes. In other words, M satisfies the IMH# if M is #-generated and whenever a first-order sentence holds in a #-generated outer model of M, it also holds in a definable inner model of M.

Theorem 4 *There are members of the hyperuniverse which satisfy IMH#.*

For a proof see Friedman and Honzik (2016). The attraction of IMH# is that it captures aspects of both vertical and horizontal maximality simultaneously.

We also mention some strengthenings of the principles given above. An *absolute parameter* is a set p which is uniformly definable over all outer models of V which 'respect p', i.e. which preserve cardinals up to and including the cardinality of the transitive closure of p. The SIMH (Strong IMH) is the IMH for sentences with absolute parameters relative to outer models which respect them: if a sentence with absolute parameters holds in an outer model which respects those parameters then it holds in a definable inner model.

A related principle is the CPIMH (Cardinal Preserving IMH). A *cardinal-absolute parameter* is a set p which is uniformly definable over all cardinal-preserving extensions of V. Then CPIMH asserts that if a sentence with cardinal-preserving parameters holds in a cardinal-preserving outer model of V it also holds in a definable inner model of V.

Restricting to #-generated universes yields corresponding principles SIMH# and CPIMH#.

We do not know whether there are elements of $\mathbb{H}$ satisfying SIMH, CPIMH or their #-versions, but it is reasonable to conjecture that they do.[13] We have:

Theorem 5 (see Friedman 2006)

(a) In all universes satisfying IMH, PD is false, and there are no large cardinals.

(b) All universes which satisfy SIMH, CPIMH or their #-versions also satisfy $\neg$*CH.*

[13] In particular, there are universes which obey them restricted to the parameter ω_1.

Thus maximality principles emanating from the Hyperuniverse Programme do indeed have striking first-order consequences.

10.6 New Axioms as $\mathbb{H}$-Axioms

10.6.1 The Nature of $\mathbb{H}$-Axioms

As we said at the beginning, we do not want to advocate any specific first-order new axiom in this paper, but rather present an alternative conceptual framework whereby higher-order statements are indeed new axioms, which also happen to have important first-order consequences. The framework we have presented, in particular the mathematical results detailed in the previous section, lend support to the following conclusion: members of $\mathbb{H}$ satisfying maximality principles have remarkable properties, e.g. in all countable transitive models satisfying IMH, PD is false and in all of them satisfying CPIMH#, CH is false.

Now, let us focus our attention for a moment on ¬PD and ¬CH. These first-order set-theoretic statements are consequences of new axioms that:

(1) hold in 'local' areas of $\mathbb{H}$
(2) are expressed in terms of intrinisically motivated maximality principles as, respectively, IMH and CPIMH#.

By virtue of this, we label IMH and CPIMH# $\mathbb{H}$-axioms, insofar as they hold in specific portions of $\mathbb{H}$ and are intrinsically motivated on the grounds of the [MIC] and [MaxExt].

Again, it is important to emphasise on what grounds our claim can be made: using V-logic, we can characterise the relationship between maximality principles and their consequences as mirroring that between higher-order properties of V and first-order truths in members of $\mathbb{H}$. In particular, in the Hyperuniverse Programme higher-order properties of V are, in a sense, turned into $\mathbb{H}$-axioms, properties of members of $\mathbb{H}$ expressible through (first-order) quantification over $\mathbb{H}$.

Furthermore, we also claim that ¬PD would be, in accordance with our conceptual presuppositions, an intrinsically motivated new set-theoretic truth insofar as IMH is an intrinsically motivated maximality principle.

Of course there are members of $\mathbb{H}$ which do not satisfy the IMH. Consequently, ¬PD is a statement holding only in a portion of $\mathbb{H}$, something which accounts for our idea that $\mathbb{H}$-axioms are 'local' axioms. This is inevitable if one wishes to be conceptually faithful to the multiverse phenomenon.

However, there is a global corrective to this 'pluralistic' view. The programme strives for the identification of an 'optimal' maximality principle ($\mathbb{H}$-axiom). Now, suppose that P were such a principle; we would then exclude any member of $\mathbb{H}$ which would not satisfy P and therefore P could be taken to be the 'new' $\mathbb{H}$-axiom we are searching for, derivable from the maximal iterative conception and with intrinsically justified first-order consequences.

It could be objected that viewing new truths as 'consequences' of more general principles ($\mathbb{H}$-axio) implies that one accepts these 'consequences' without understanding their 'content', in particular whether they are 'intuitively true' and this would distance our methodology from a genuine search for 'meaningful' additions to ZFC. However, the methodology envisaged here precisely aims to provide an alternative notion of 'intuitively true' as based on the acceptance of the intuitive truth of maximality principles concerning V. Therefore, in our view, the 'meaningfulness' of the consequences of a maximality principle is guaranteed by the meaningfulness of the principle itself.

10.6.2 Alternative Approaches

Finally, we go back, again, to the issue we started with at the beginning of this paper: what new axioms should be. First of all, we will try to dispel one main worry about the methodology described, namely, that it could imply that all new axiom candidates other than $\mathbb{H}$-axioms should automatically fail to be viewed as plausible new axioms and, what is worse, as lacking any evidence in favour of their acceptance.

This would be a gross misrepresentation of our perspective. In the previous subsection, when we regarded $\neg$PD or $\neg$CH as consequences of new axioms, our aim was not to make a general argument in favour of the rejection of PD or CH. At the same time, nowhere in this paper have we suggested that the 'current' new axioms should *all* be rejected: the proof of this is that, again, PD, CH or their negations have already been subjected to extensive mathematical investigations as new axioms, and, in this respect, our programme has nothing new to add.

What we have tried to establish here is that, *if* our evidential framework is preferable to others, then there are reasons to think that PD might be rejected precisely on its grounds.

Leaving aside our framework for a moment, it is maybe appropriate to make a brief digression on the status of PD. Over the years, PD has been celebrated as a new axiom for which there is a significant body of evidence.[14] In particular, two aspects are almost invariably highlighted: (1) PD is successful, because it makes the theory of sets of reals up to and including the projective sets behave well (under PD, all projective sets of reals are Lebesgue measurable and every uncountable projective set of reals has a perfect subset, which means that CH cannot be projectively refuted); (2) PD remarkably connects two apparently distant areas of set theory, descriptive set theory and the theory of large cardinals, as it was proved that the existence of Woodin cardinals and PD have the same consistency strength.

---

[14] For the full case for axioms of definable determinacy, such as PD, see, e.g., Woodin (2001), Martin (1998), Koellner (2006).

However, arguments in favour of PD are mostly extrinsic and are based on the fact that it follows from large cardinals or from set-generic absoluteness principles, but justifications are lacking for both the existence of large cardinals and for a form of absoluteness which imposes an artificial restriction to set-forcings.[15] Also, advocates of PD often claim that truth is taken to be based solely on current set-theoretic practice, ignoring what is relevant for mathematics outside of set theory or for the maximal iterative conception. So arguing that PD can be inferred from current set-theoretic practice may be insufficient for claiming its truth.[16]

Now, returning to our main topic, why should all other proposed definitions of what a 'new axiom' should be like be replaced by ours? Because other approaches may be fraught with insurmountable difficulties. For instance, consider the following three alternatives:

A new axiom should be a first-order statement true of the concept of set. As we have seen, true of the concept of set means true of the [MIC], but there might be quite a few set-theoretic statements for which such criterion cannot apply. For instance, is the Axiom of Choice true of the [MIC]? How about the Axiom of Determinacy? Even if such criterion is applicable, there might be cases where one intrinsically motivated first-order axiom may contradict another enjoying the same status.

A new axiom should be a first-order statement, not intrinsically, but rather extrinsically justified. Many new axioms such as forcing axioms or PD and, in general, definable determinacy axioms, have a lot of strong extrinsic support. However, this fact may not be sufficient and, in fact, too limited. For instance, in the Appendix (Sect. "Appendix: Absoluteness Axioms"), we present arguments showing that extrinsically supported absoluteness axioms may be inadequate.

A new axiom is an axiom which is 'practically' confirmed, that is, verified empirically in specific areas of set theory'. This is a refinement of the statement above. However, the definition is still problematic, as the notion of an axiom's being 'practically confirmed' is obscure and would require clarification.[17]

[15]On this, see Appendix of the present paper.

[16]To be fair, advocacy of PD along an alternative, intrinsic-evidence-based line of thought, has also been made. See, for instance, Hauser (2002): 'But aside from extrinsic evidence, there are other reasons to regard *PD* as the *correct* axiom for the projective sets. With the progress made in the theory of canonical models for large cardinals it has become clear that *PD* is implied by and is in fact equivalent to a vast number of *prima facie* unrelated combinatorial principles including large-cardinal axioms. Still this may not establish their intrinsic necessity because the relevant large-cardinal axioms at present do not enjoy the same kind of intrinsic plausibility as for example Mahlo cardinals. However, the intrinsic necessity of an axiom need not be immediate and could depend on the discovery of additional facts' (p. 274). Of course, at present it is not clear what 'intrinsic' facts would add to the defensibility of PD and whether they will ultimately be discovered.

[17]For further details on these different approaches, see, respectively: (1) on the strength and value of extrinsic justifications, Maddy (1996, 1997), Koellner (2006), Martin (1998); (2) on second-order logic and set theory, Shapiro (1991), and Jané (2005a); (3) on the quasi-empirical view, again, Koellner (2006), or Hauser (2002).

We do not know whether the above procedure to identify and justify $\mathbb{H}$-axioms and the notion of $\mathbb{H}$-axiom itself will become standard. It does seem to us that our proposal responds better to the conceptual difficulties of the aforementioned alternative approaches. In particular, after the substantial demise of 'Gödel's programme', the search for new intrinsically motivated new axioms is at a loss within all other current research programmes. The reasons have been amply considered above, especially in our introductory remarks: the notion of set-theoretic truth falls short of a unique characterisation, if it is to reflect a unique realm of objects, in particular as a consequence of the existence of the multiverse, and it does not seem that this situation can be easily repaired, unless one adopts higher-order principles motivated by the concept of set.

10.7 Concluding Summary

In this paper, we have shown how the search for new axioms is carried out within the Hyperuniverse Programme. The methodology devised is motivated by the existence of three concurrent phenomena: (1) the set-theoretic multiverse; (2) the availability of higher-order principles describing forms of maximality of V in line with the [MIC], that is, $\mathbb{H}$-axioms; (3) a demonstrable link between such maximality principles and countable transitive models.

Maximality principles, that specify different notions of the maximality of V, also have, through the use of V-logic, robust consequences in countable transitive models. Obviously, different maximality principles may have different first-order consequences. So, the main shortcoming of this conception is that it is not sufficient to fix set-theoretic indeterminacy *uniquely*. However, we believe that the further development of the programme may establish the existence of an 'optimal' maximality criterion, which, in turn, may lead to the acceptance of one single, intrinsically justified collection of first-order statements to be added to ZFC.

The project is open to further generalisations and developments. New maximality principles will come out, helping us to identify further universes where certain set-theoretic statements do or do not hold. A more careful description of V, of different types of universes in $\mathbb{H}$, and axioms therein, may, therefore, be on its way.

Appendix: Absoluteness Axioms

In recent years, considerable attention has been paid by set-theorists to what we may call the *absoluteness programme*. The main goal of this programme is to foster suitable mathematical strategies and principles (absoluteness axioms) to 'induce' the absoluteness of certain set-theoretic statements across an appropriately selected collection of models (or set-theoretic multiverse).[18]

Although absoluteness axioms have received a lot of extrinsic support in recent years,[19] here we want to present evidence that no new first-order absoluteness axiom has good prospects to be viewed as a plausible axiom candidate extending ZFC. As far as the 'extrinsic' value of these axioms is concerned, the reasons for this claim are *structural*, that is, refer to internal features of the absoluteness phenomenon and do not depend upon the nature and the content of the axiom under consideration.

We now explain why this is so.

Recall the Lévy hierarchy of logical formulas: one starts with Δ_0-sentences, those with only *bounded* quantifiers. Σ_1- and Π_1-sentences contain, respectively, one block of existential or one block of universal quantifiers followed by bounded quantifiers and, in general, $\Sigma_{k+1} = \exists x_1 x_2 \ldots x_n \Pi_k$ and $\Pi_{k+1} = \forall x_1 x_2 \ldots x_n \Sigma_k$. Also, recall that $H(\kappa)$ denotes the union of all transitive sets of size less than κ.[20] The Σ_n-theory of $H(\kappa)$ is the set of Σ_n-sentences true in $H(\kappa)$.

Definition 3 We say that $M \sqsubseteq N$ if $M \subseteq N$ are transitive models of ZFC with the same ordinals.

Now, there exist trivial forms of absoluteness. For instance, as is known, if $M \sqsubseteq N$, where M and N are models of ZFC, the theory of $H(\omega)$ is the same in M and N. Going one level higher in the hierarchy of $H(\kappa)$, one finds the following seminal result due to Lévy and Shoenfield:

Theorem 6 *If $M \sqsubseteq N$ are models of ZFC, then the Σ_1-theory of $H(\omega_1)$ is the same in M and N.*

Now, what about the Σ_2-theory of $H(\omega_1)$? Climbing up the scale of complexity of set-theoretic sentences, absoluteness comes to a halt:

Theorem 7 *There are models $M \sqsubseteq N$ of ZFC such that the Σ_2-theory of $H(\omega_1)$ is not the same in M and N.*

Proof The statement "there is a nonconstructible real" is a Σ_2 property of $H(\omega_1)$. Take N to satisfy this and M to be L^N. □

This negative result may be circumvented via a two-step strategy: the first step consists in restricting the $\sqsubseteq$-relation in a suitable way. Consider the following definition:

Definition 4 $M \sqsubseteq^{set-generic} N$ iff N is a set-generic extension of M.

---

[18]Given a formula ϕ and transitive models M and N, we say that ϕ is absolute between M and N iff $\phi^M(x_1, x_2, \ldots, x_n) \leftrightarrow \phi^N(x_1, x_2, \ldots, x_n)$.

[19]For an introductory overview of some of these see Bagaria (2000).

[20]That is, the union of all sets whose *transitive closure* has cardinality less then κ.

Theorem 8 (Bukovsky[21]) *$M \sqsubseteq^{set-generic} N$ iff $M \sqsubseteq N$ and for some cardinal κ of M every function in N on a set in M into M is contained in a multi-valued function in M with fewer than κ values for each argument.*

One further refinement of this definition leads to the following notion:

Definition 5 $M \sqsubseteq^{stationary-preserving-set-generic} N$ iff N is a set-generic extension of M and any subset of ω_1^M which is stationary in M is also stationary in N.

In other terms, by restricting the $\sqsubseteq$-relation to, respectively, $\sqsubseteq^{set-generic}$ or, on the other hand, $\sqsubseteq^{stationary-preserving-set-generic}$ one only takes into account *generic extensions* of models obtained through *set-forcing* or *stationary-preserving set-forcing*.

The second step in the strategy consists in considering certain extensions of ZFC, say, ZFC + Ax., and then replacing the multiverse $\mathbb{M}^{ZFC}$ by the multiverse $\mathbb{M}^{ZFC+Ax.}$ associated to the stronger system ZFC + Ax.

Using this two-step strategy, Woodin and Viale have obtained results which are, no doubt, of mathematical significance,[22] but, with respect to our foundational project, their results present some crucial shortcomings: (1) the axioms they consider, such as the existence of class-many Woodin cardinals, are not justified *intrinsically*; (2) the restriction of the $\sqsubseteq$-relation to set-generic extensions is unwarranted in view of our definition of $\mathbb{H}$. Furthermore, even leaving these issues aside, it is not clear how far the programme they have been carrying out can be extended, and with what results. We will come to this in a moment.

Alternatively, one could employ only the second step of the above strategy, by supporting the acceptance of axioms such as $V = L$. Gödel's work yields:

Theorem 11 *If $M \sqsubseteq N$ are models of $ZFC + V = L$, then $M = N$.*

However, promising though this strategy may seem, it reveals the same shortcoming as before, insofar as it is hinged upon the acceptance of a mathematical principle, $V = L$, which does not possess a sufficient degree of intrinsic motivation in view of our notion of 'intrinsicness' expounded in Sect. 10.5.

As said, in fact, there is strong evidence that the second step in the above two-step strategy, that of extending ZFC to a stronger first-order theory to obtain greater absoluteness, is doomed to failure. Consider the following:

---

[21]See Bukovsky (1973).

[22]See, in particular, Woodin (2001) and Viale (2016). Among other things, Woodin proved the following:
Theorem 9 *If $M \sqsubseteq^{set-generic} N$ are models of ZFC+ large cardinals + CH, then the Σ_1-theory of $H(\omega_2)$ (with parameter ω_1) is the same in M and N.*
Viale has recently proved:
Theorem 10 *If $M \sqsubseteq^{stationary-preserving-set-generic} N$ are models of ZFC + large cardinals+MM^{+++}, then the theory of $H(\omega_2)$ is the same in M and N.*

Theorem 12 *Suppose T is a first-order theory, compatible with the following two statements:*

(1) *the class $\{\alpha : \alpha measurable\}$ is stationary;*
(2) *the class $\{\alpha : V_\alpha \prec_{\Sigma_\omega} V\}$ is unbounded.*

Then, $\Sigma_2(H(\omega_1))$-absoluteness fails for models of T: there are models $M \sqsubseteq N$ of T such that the Σ_2-theory of $H(\omega_1)^M \neq \Sigma_2$-theory of $H(\omega_1)^N$. If T consists of only a finite set of axioms then (2) above is not needed.

Sketch of Proof. The hypotheses imply that there is a model V of ZFC with a largest measurable κ such that T holds in V_κ. Now iterate the measure on κ through the ordinals, resulting in a model N. In N, there is a model V_0 like V but only satisfying KP, with an iterable top measure. Again iterate the top measure through the ordinals to form an inner model M. Then $M \sqsubseteq N$ are both models of T but by choosing V_0 minimally we can arrange that in M there is no iterable model P of KP with a top measurable κ_0 such that T holds in the V_{κ_0} of P. This $\Pi_2(H_{\omega_1})$ sentence fails in N and this gives the asserted failure of absoluteness. □

The theorem asserts that any first-order theory which is compatible with the existence of *measurable cardinals* (in fact, a *stationary class* of measurable cardinals) fails to ensure $\Sigma_2(H(\omega_1))$ absoluteness for its models. This is very strong evidence against the use of first-order axioms for obtaining convincing absoluteness principles for $\Sigma_2(H(\omega_1))$ statements.

To summarise, there is a network of results which seem to show that, through the adoption of absoluteness axioms, one can find new set-theoretic truth, by extending the absoluteness of set-theoretic statements to levels of increasing first-order complexity. However, first of all, none of the axioms adopted or used in the programme seems to be intrinsically motivated. Secondly, there is also some evidence that such an extension collides with the existence of measurable cardinals. As a consequence, one appears to be forced to artificial restrictions of the multiverse to only certain models or of the notion of absoluteness itself.

Consequently, we come to the following conclusion: no first-order *absoluteness axiom* has good prospects of being accepted as a new axiom on the grounds of both intrinsic or extrinsic justifications.

In our opinion, if one wants to spell out a plausible conception of 'truth in the multiverse', one has to proceed in the alternative way we propose, through the use of higher-order principles.

References

Antos, C., Friedman, S.-D., Honzik, R., & Ternullo, C. (2015). Multiverse conceptions in set theory. *Synthese*, *192*(8), 2463–2488.

Arrigoni, T., & Friedman, S. (2012). Foundational implications of the inner model hypothesis. *Annals of Pure and Applied Logic*, *163*, 1360–1366.

Arrigoni, T., & Friedman, S. (2013). The hyperuniverse program. *Bulletin of Symbolic Logic*, *19*(1), 77–96.

Bagaria, J. (2000). Bounded forcing axioms and generic absoluteness. *Archives for Mathematical Logic*, *39*, 393–401.

Barwise, J. (1975). *Admissible sets and structures*. Berlin: Springer.

Benacerraf, P., & Putnam, H. (Eds.). (1983). *Philosophy of mathematics. Selected readings*. Cambridge: Cambridge University Press.

Bernays, P. (1935). Sur le platonisme dans les mathématiques. *L'Enseignement Mathématique*, *34*, 52–69.

Bukovsky, L. (1973). Characterization of generic extensions of models of set theory. *Fundamenta Mathematicae*, *83*, 35–46.

Ewald, W. (Ed.). (1996). *From Kant to Hilbert: A source book in the foundations of mathematics* (Vol. II). Oxford: Oxford University Press.

Friedman, S. (2006). Internal consistency and the inner model hypothesis. *Bulletin of Symbolic Logic*, *12*(4), 591–600.

Friedman, S. (n.d.). The philosophy and mathematics of set-theoretic truth. In *Proceedings of the 2014 Chiemsee summer school 'proof, truth and computation'. IF CoLog Journal of Logics and their Applications*.

Friedman, S., & Honzik, R. (2016). On strong forms of reflection in set theory. *Mathematical Logic Quarterly, 62*(1–2), 52–58.

Friedman, S., Welch, P., & Woodin, W. H. (2008). On the consistency strength of the inner model hypothesis. *Journal of Symbolic Logic*, *73*(2), 391–400.

Gödel, K. (1947). What is Cantor's continuum problem? *American Mathematical Monthly*, *54*, 515–525.

Hallett, M. (1984). *Cantorian set theory and limitation of size*. Oxford: Clarendon Press.

Hauser, K. (2002). Is the continuum problem inherently vague? *Philosophia Mathematica*, *10*, 257–285.

Hellman, G. (1989). *Mathematics without numbers: Towards a modal-structural interpretation*. Oxford: Clarendon Press.

Horsten, L., & Welch, P. (2013). Absolute infinity. Unpublished.

Jané, I. (1995). The role of the absolute infinite in Cantor's conception of set. *Erkenntnis*, *42*, 375–402.

Jané, I. (2005a). Higher-order logic reconsidered. In S. Shapiro (Ed.), *Oxford handbook of philosophy of mathematics* (pp. 747–774). Oxford: Oxford University Press.

Jané, I. (2005b). The iterative conception of sets from a Cantorian perspective. In P. Hájek, D. Westertål, & L. Valdés-Villanueva (Eds.), *Logic, methodology and philosophy of science. Proceedings of the twelfth international congress* (pp. 373–393). London: King's College Publications.

Koellner, P. (2006). On the question of absolute undecidability. *Philosophia Mathematica*, *14*(2), 153–188.

Koellner, P. (2009). On reflection principles. *Annals of Pure and Applied Logic*, *157*(2–3), 206–219.

Linnebo, Ø. (2013). The potential hierarchy of sets. *Review of Symbolic Logic*, *6*(2), 205–228.

Maddy, P. (1988a). Believing the axioms. *I. Bulletin of Symbolic Logic*, *53*(2), 481–511.

Maddy, P. (1988b). Believing the axioms. *II. Bulletin of Symbolic Logic*, *53*(3), 736–764.

Maddy, P. (1996). Set-theoretic naturalism. *Bulletin of Symbolic Logic*, *61*(2), 490–514.

Maddy, P. (1997). *Naturalism in mathematics*. Oxford: Oxford University Press.

Martin, D. (1998). Mathematical evidence. In H. G. Dales & G. Oliveri (Eds.), *Truth in mathematics* (pp. 215–231). Oxford: Clarendon Press.

Parsons, C. (1983). *Mathematics in philosophy*. Ithaca: Cornell University Press.

Potter, M. (2004). *Set theory and its philosophy*. Oxford: Oxford University Press.

Shapiro, S. (1991). *Foundations without foundationalism: A case for second-order logic*. Oxford: Oxford University Press.

Tait, W. W. (1998a). Foundations of set theory. In H. G. Dales & G. Oliveri (Eds.), *Truth in mathematics* (pp. 273–290). Oxford: Oxford University Press.

Tait, W. W. (1998b). Zermelo on the concept of set and reflection principles. In M. Schirn (Ed.), *Philosophy of mathematics today* (pp. 469–483). Oxford: Clarendon Press.

Viale, M. (2016). Category forcings, MM^{+++} and generic absoluteness for the theory of strong forcing axioms. *Journal of the American Mathematical Society, 29*(3), 675–728.

Wang, H. (1974). *From mathematics to philosophy*. London: Routledge & Kegan Paul.

Wang, H. (1996). *A logical journey*. Cambridge: MIT Press.

Woodin, W. H. (2001). The continuum hypothesis. *Notices of the American Mathematical Society*, Part 1: *48*(6), 567–76; Part 2: *48*(7), 681–90.

Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche: neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae*, *16*, 29–47.

Chapter 11
Multiversism and Concepts of Set: How Much Relativism Is Acceptable?

Neil Barton

Abstract Multiverse Views in set theory advocate the claim that there are many universes of sets, no-one of which is canonical, and have risen to prominence over the last few years. One motivating factor is that such positions are often argued to account very elegantly for technical practice. While there is much discussion of the technical aspects of these views, in this paper I analyse a radical form of Multiversism on largely philosophical grounds. Of particular importance will be an account of *reference* on the Multiversist conception, and the *relativism* that it implies. I argue that analysis of this central issue in the Philosophy of Mathematics indicates that Radical Multiversism must be *algebraic*, and cannot be viewed as an attempt to provide an account of reference without a softening of the position.

Keywords Philosophy of mathematics · Set theory · Foundations of mathematics · Multiverse

11.1 Introduction

The development of set theory since the 1930s has been partly shaped by a fascinating phenomenon: independence results. Since Gödel's proof of the consistency of the Axiom of Choice and Continuum Hypothesis with Zermelo–Fraenkel set theory, various model-theoretic techniques (notably the method of forcing developed by Cohen) have allowed us to show that for many sentences of set theory ϕ, it is neither the case that $ZFC \vdash \phi$ nor $ZFC \vdash \neg\phi$. The case is particularly acute in set theory, as the kinds of statements that turn out to be independent from ZFC are extremely natural questions, rather than gerrymandered statements of metalogic (such as the Gödel sentence for Peano Arithmetic). While the most infamous example is probably

N. Barton (✉)
Department of Philosophy, Birkbeck College, University of London,
Malet Street, London, UK
e-mail: nbarto02@mail.bbk.ac.uk

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_11

the Continuum Hypothesis, the phenomenon is visible with respect to a wide range of mathematical entities. So Hamkins writes:

> A large part of set theory over the past half-century has been about constructing as many different models of set theory as possible, often to exhibit precise features or to have specific relationships with other models. Would you like to live in a universe where *CH* holds, but $\diamondsuit$ fails? Or where $2^{\aleph_n} = \aleph_{n+2}$ for every natural number n? Would you like to have rigid Suslin trees? Would you like every Aronszajn tree to be special? Do you want a weakly compact cardinal κ for which $\diamondsuit_\kappa(REG)$ fails? Set theorists build models to order (Hamkins 2012a, p. 417)

There have been (broadly speaking) two reactions to this phenomenon. On the one hand, some regard these results as indicative of a failure of bivalence in set theory. Others, however, prefer to maintain that these sentences nonetheless have a definite truth value.

In this paper I am primarily concerned with one view that regards some sentences of set theory as non-bivalent. This is Joel David Hamkins' recently proposed and technically elegant 'Multiverse View', given in Hamkins (2012a). I shall argue that Hamkins' position may be interpreted in one of two ways, either as giving an account of *ontology* and *reference* (with similarities to ideas presented in Balaguer (1998)), or alternatively as providing a framework for understanding set theory *algebraically*. Taking Hamkins' view in the former way, I suggest, leads to a referential regress: on a given occasion of purported set-theoretic reference there is a vicious non-well-founded dependency chain. A similar problem, I argue, can be pressed against Hamkins concerning various metalogical notions. Analysis of these problems reveals that if the Hamkinsian Multiversist wishes to have her view interpreted ontologically, she[1] has to de-radicalise some of her theses. My strategy is as follows:

In Sect. 11.2 I outline the version of Multiversism under consideration (given in Hamkins 2012a) and some of its theoretical motivations. I note that there are at least two ways of interpreting his view: either as providing a philosophical view of ontology, or specifying an algebraic framework for the practice of set theory. On the former interpretation, I note that the method by which we *refer* to the subject matter of mathematics is usually through *description*, and is often argued to solve a problem present in Benacerraf (1973). In Sect. 11.3 I point out that this way of interpreting Hamkins' Multiverse View leads to a very strong form of *relativism*. I then use this to develop a regress, one which I argue is vicious. Further analysis reveals that if we wish to refer to mathematical objects in a philosophically satisfactory manner, a core of concepts must be taken to be determinately understood. Section 11.4 then examines some consequences of this relativism for the study of metalogic. It is argued that the inability of the Multiversist to provide a characterisation of finitude appears deeply problematic and is further evidence that if we are to maintain an account of reference, then certain mathematical notions must be kept absolute. In Sect. 11.5 I argue that the algebraic interpretation does not suffer from these difficulties, but note that it does not provide an answer to Benacerraf's problem.

---

[1]In order to disambiguate, throughout this paper I speak as if the Hamkinsian Multiversist is female and her opponents are male (with the exception of Hamkins himself).

It is concluded that the Hamkinsian Multiversist must either soften her position to include some concepts as determinately understood, or alternatively take her view to be one not concerned with ontology and reference.

11.2 The Radical Multiverse View

To see from where the Multiverse View gains its motivation, it will be useful to first examine its polar opposite. Pre-theoretically, once we have accepted the Iterative Conception of Set as the conceptual underpinning of *ZFC* set theory[2] it is natural to hold the following view:

> [Universism] There exists a single, unique, maximal, determinate universe of sets, to which we may refer precisely using our set-theoretic concepts (all of which, in turn are determinate)[3]

Universism thus ensures that every statement of set theory is determinately true or false, we have a determinate ontology and concepts that we can use to refer precisely to said ontology. The Multiversist takes it, however, that there are good reasons to reject Universism. In particular I will be concerned with a very radical form of Multiversism given in Hamkins (2012b). A main argument of this paper will be that the position can be interpreted in different ways. For this reason, we begin with a coarse grained characterisation of his view and refine it later.

Explaining his philosophical standpoint, Hamkins writes:

> As a result [of the independence phenomenon], the fundamental objects of study in set theory have become the models of set theory, and set theorists move with agility from one model to another (Hamkins 2012b, p. 418)

In order to arrive at a first approximation, I shall take this as an initial statement of Hamkins' view. He claims the fundamental objects of set-theoretic study are the *models* of different set theories. This characterisation immediately raises the following question: what theory should these models satisfy?

For the purposes of this paper, we shall take the models to satisfy first-order *ZFC*. There are two reasons for this choice. First, while Hamkins is potentially open to the consideration of different multiverses (such as the multiverse of second-order arithmetic), it is *ZFC* set theory with which he is primarily interested. Second, the theory must be first-order as Hamkins holds that our interpretation of second-order vari-

---

[2]We should be mindful here of the fact that the idea of the Iterative Conception underpinning *ZFC* set theory is quite controversial in itself. There is an extensive literature on the topic, for a small selection see Maddy (1988) and Potter (2004). I do not address the question further here, and take it as assumed that the Iterative Conception is the justificatory resource to which the Universist appeals.

[3]It should be noted that there is the scope to hold that there is just one Universe of sets but that it is indeterminate in some sense (say because, some of its *properties* are indeterminate). For just such a view, see Feferman (2011).

ables requires a background set theory. For this reason, he holds that indeterminacy in first-order ZFC would transfer to the second-order setting.[4]

Thus, we arrive at the following initial characterisation of Hamkins' position:

> [Hamkinsian Multiversism] There is no one universe of sets, but rather many. A universe of sets is simply a model of first-order ZFC.

A remark on terminology is important here. The Multiversism in question must be specified to be Hamkinsian, as there are many Multiverse views, each of which takes different concepts to be determinate. For example, a Zermelian form of Multiversism normally holds questions of 'height' (i.e. the ordinals that exist) to be indeterminate, but questions of 'width' (i.e. the subsets formed at $V_{\alpha+1}$ given some V_α) to be determinate.[5] On such a picture the universes are the natural models of second-order[6] ZFC_2. Hence, questions concerning the low levels of the hierarchy (such as CH) are determinate: the relevant V_κ is the same in every natural model of ZFC_2. Many questions of height (such as the existence of an inaccessible cardinal) turn out to be indeterminate, as there are usually some models of ZFC_2 in which there are no cardinals of the required variety (for example, the smallest model of ZFC_2 does not contain an inaccessible). Conversely, we might hold a version of Multiversism on which questions of height are determinate, but questions of width (for example the Continuum Hypothesis) are indeterminate (such a view is advocated by Steel 2014). Here, our universes would be the models which contain all the ordinals, including their inner models and (set) forcing extensions. Hamkinsian Multiversism represents a very extreme position in this regard, a universe on his view is simply any structure that satisfies first-order ZFC.

In order to understand what is at stake for the Hamkinsian Multiversist, I explain three motivations that have been given for the view: (1) the generality of forcing constructions, (2) the avoidance of arbitrariness, and (3) the possibility of additional mathematical insight.

11.2.1 The Generality of Forcing

Universism, it might be argued, faces a substantial challenge in the form of the independence results. It is not simply the fact that there are statements that are independent (a fact that, though it requires explanation, may be dismissed as a side issue about the paucity of our proof-theoretic framework), but rather the *manner* in which these results are proved.

---

[4] This fact will turn out to be important for assessing Hamkins' view, and is discussed in more detail in Sect. 3.

[5] The idea has its conceptual roots in Zermelo (1930), but see also Isaacson (2011).

[6] This is only true, of course, if we are allowed to use the 'full' semantics for ZFC_2. See Meadows (2013) and Koellner (2013) for discussion of the surrounding philosophical issues.

The issue is the following. Set-theoretic practice is replete with model-theoretic techniques that, given a particular model of set theory $\mathfrak{M}$, 'add' sets not already in $\mathfrak{M}$ to $\mathfrak{M}$. The clearest technique with which this is visible is *forcing*. In order to force over a particular model of set theory $\mathfrak{M}$ we begin with a partial order $\mathbb{P} = \langle P, <_{\mathbb{P}}, \mathbb{1}_{\mathbb{P}} \rangle \in \mathfrak{M}$. We then choose a filter G on $\mathbb{P}$ that intersects every dense subset D of P that lies in $\mathfrak{M}$. Next, we add G to $\mathfrak{M}$ and close under set-theoretic operations definable in $\mathfrak{M}$. The end result is a new model $\mathfrak{M}[G]$ that (given the correct choice of partial order) satisfies the axioms and the appropriate sentence for the independence proof. Importantly, however, $G \notin \mathfrak{M}$ for non-trivial forcing constructions.[7] The challenge for the Universist lies in the fact that often set theorists will force using 'V' to denote the ground model. Normally, 'V' denotes the universe of *all* sets. Since the set added (call it 'G') cannot be in the model over which we are forcing, the Universist faces a problem. For, it appears that they should say that the statement "there is a forcing extension $V[G]$ such that..." is true: many forcing constructions make just this claim and the mathematics in question appears perfectly rigorous and consistent. However, while G is a set, G cannot be in V by construction, but V is supposed to be all the sets there are. It looks as though something has to give.

At this point, it should be noted that the Universist has several quite robust options involving complex simulations of the statements in question. He could, for example, use the Reflection Theorem in conjunction with the Downward Löwenheim–Skolem Theorem to obtain a countable structure over which he can force.[8] Alternatively, he could interpret the forcing *via* Boolean-valued models (or if she wishes to interpret the reasoning in a two-valued manner, use a Boolean ultrapower and quotient structure[9]). Still further, he could analyse the problematic statements syntactically using the forcing relation.

All this is satisfactory as far as it goes, and allows the Universist to interpret forcing talk where the symbol 'V' is used. The Multiversist rejoinder, however, is that it seems arbitrary to insist that V does not have a forcing extension where the relevant notions seem perfectly well defined. The issues here are subtle and complex, and I defer consideration of Universist interpretations of forcing to different work. Suffice to note for current purposes that the Universist has to provide a complex and detailed paraphrase strategy in interpreting this set-theoretic discourse.

This is precisely what is not required under Hamkinsian Multiversism, however. For her, the account of set-theoretic forcing is as follows. We start with a particular set-theoretic background, that we take to be 'all' the sets and dub this model 'V'. Then, when we perform the sorts of model-theoretic constructions that necessitate the addition of sets to the model, we simply move from V to a different (equally legitimate) universe of sets that may (and usually does) satisfy very different sen-

[7]For details see Kunen (1980) and Jech (2002).

[8]In particular, this strategy seems attractive for the reason that often use of the term 'V' is patently an abuse of notation, designed to underscore the fact that *any* countable transitive model will suffice for the construction. See Koellner (2013) for discussion of some of these issues.

[9]For details, see Hamkins and Seabold (2012).

tences. We do not have to provide any awkward paraphrase, rather, we can hold that when a set theorist asserts statements about 'the universe' V and moves to a forcing extension $V[G]$, she begins by considering some particular universe and then shifts reference to a different universe (i.e. the forcing extension). Where the Universist has to use a paraphrase strategy to interpret the claim "V has a forcing extension $V[G]$ such that...", the Multiversist can simply assert that reference is entirely transparent: she begins by referring to a perfectly real and legitimate universe, and then simply shifts reference to an equally real and legitimate universe. In short, the Hamkinsian Multiversist contends that she avoids insisting in an ad hoc fashion that V has no forcing extension, she can hold any construction involving the term 'V' to have perfectly good and transparent ontological reference.

11.2.2 The Avoidance of Arbitrariness

The forcing construction provides a clear example of where Universism has questions to answer in interpreting set-theoretic discourse. However, the problem is indicative of a wider issue in set theory:

> the most prominent phenomenon in set theory has been the discovery of a shocking diversity of set-theoretic possibilities. Our most powerful set-theoretic tools, such as forcing, ultrapowers, and canonical inner models, are most naturally and directly understood as method of constructing alternative set-theoretic universes. (Hamkins 2012b, p. 418)

The point here is fairly simple: while forcing presents the most obvious conflict with Universism (in virtue of adding sets to the ground model) it nonetheless remains the case that there are *many* constructions that can be naturally understood from a Hamkinsian Multiversist perspective. Any particular *ZFC*-preserving construction can be interpreted as movement within the Multiverse. The Universist, in virtue of holding that there is a unique privileged Universe of set theory, has to explain why the seemingly perfectly natural set-theoretic constructions produce nothing more than model-theoretic representations within V rather than *bona fide* universes. Again, the issues here are subtle and a full consideration is outside the scope of this paper.[10] However, we should note that the Hamkinsian sees it as a substantial theoretical virtue that she can interpret set theorists as concerned simply with other models that are fully legitimate universes in their own right, rather than having to provide a philosophical analysis of the difference between V and these other structures.

11.2.3 Additional Mathematical Insight

The final reason I shall mention here is that (Hamkins claims) the Universist runs the risk of missing mathematical insights. He argues as follows:

[10] Again, some of these issues are given consideration in Koellner (2013).

> Such a perspective may be entirely self-consistent, and I am not arguing that the universe view is incoherent, but rather, my point is that if one regards all outer models of the universe as merely simulated inside it via complex formalisms, one may miss out on insights that could arise from the simpler philosophical attitude taking them as fully real (Hamkins 2012a, p. 426)

The key point here is the following: model-theoretic constructions are no longer merely devices for showing relative consistency proofs (as was the case with Gödel and Cohen). Rather, such constructions are now used to show theorems about *objects* that are of interest in their own right (rather than say, a number-theoretic fact about proof codes). The emerging picture has been one in which the study of multiverses and interrelations between models (such as embedding properties between models and their forcing extensions[11]) have become of central importance in set theory. The Hamkinsian's philosophical position, she argues, facilitates this kind of thought: we can undergo the relevant constructions and study the models without worrying whether or not such talk can be 'simulated' in the context of V.

Again the issues here are philosophically subtle, and there is a burgeoning literature on the subject.[12] For the purposes of assessing the Hamkinsian on her own terms then, we may simply take the above three issues to be the motivations she would like to preserve in interpreting set theorists.

11.3 The Ontological and Algebraic Interpretations of Hamkinsian Multiversism

Let us take stock. We are now in a position where we have a coarse grained statement of a Hamkinsian position, and some considerations that motivate the view. In this section, I argue that further examination reveals two aspects to her view, one *ontological* and one *algebraic*. Later, I shall argue that the latter is the only tenable interpretation of Hamkinsian Multiversism.

11.3.1 The Ontological Interpretation

The ontological interpretation is perhaps the more intuitive way of understanding Hamkins' view. Here we regard the view as providing an account of *ontology* and *reference*. Such an interpretation is suggested by passages such as the following:

> I shall argue for a contrary [to Universism] position, the *multiverse view*, which holds that there are diverse distinct concepts of set, each instantiated in a corresponding set-theoretic universe, which exhibit diverse set-theoretic truths. Each such universe exists independently in the same Platonic sense that proponents of the universe view regard their universe to exist (Hamkins 2012a, pp. 416–417)

[11] See, for example, Foreman (2010).

[12] See, for example, Koellner (2013) and Hamkins and Seabold (2012).

So, the picture offered is one on which each universe exists fully independently in reality, much as a Universist asserts the full and mind-independent existence of her V.[13] This provides us with an account of *ontology*, the multiverse (of *ZFC*) is constituted[14] by models of first-order *ZFC*. While there may be practical or personal reasons for preferring one universe over another, no one universe is especially privileged ontologically.

Given that she now has a picture of ontology, it is incumbent upon the Hamkinsian to provide an account of epistemology. In particular, one would like to know how *reference* behaves, and its relationship to *truth* and *knowledge*.

The question of reference is important on this interpretation of Hamkinsian Multiversism. Crucially, she wishes to hold that the study of set-theoretic properties consists in their behaviour within the Multiverse, and that questions of truth are substantiated by a *real* ontology. For example, Hamkins writes:

> The multiverse view is one of higher-order realism–Platonism about universes–and I defend it as a realist position asserting the actual existence of the alternative set-theoretic universes into which our mathematical tools have allowed us to glimpse (Hamkins 2012a, p. 417)

and

> set theorists [when performing model-theoretic constructions] move with agility from one model to another (Hamkins 2012a, p. 418)

A natural way to understand claims of the above kind is through *reference*: when performing a model-theoretic construction, we start with some relevant universe V, and shift reference to a different V'. How then does this reference to different universes within the multiverse occur? Hamkins is nowhere explicit, but does say the following:

> Often the clearest way to refer to a set concept is to describe the universe of sets in which it is instantiated, and in this article I shall simply identify a set concept with the model of set theory to which it gives rise (Hamkins 2012a, p. 417)

Thus we have a picture of reference on which each concept is correlated with a model, and we refer to this concept via describing its associated model, and we refer this concept through a *description*. An immediate question is what expressive resources we may employ in providing such a description. Hamkins is very clear: any such description is first-order. The key issue here is that this account of reference to

---

[13]It should be noted that the analogy is not total. For instance, a Universist is likely to deny the existence of V as a set for reasons of paradox. On Hamkins' view, however, any particular V is a set in a taller model. Nonetheless, one can see the similarity, both assert that the subject matter of mathematics is constituted by mind-independent entities.

[14]Exactly which models of first-order *ZFC* is a subtle issue, and one I shall not consider here. Hamkins specifies (in addition to his broad picture) a list of Multiverse Axioms, designed to axiomatise movement within the Multiverse. Of note is that (within a particular V) the collection of all models of *ZFC* does *not* satisfy the Multiverse Axioms, though the collection of all computably saturated models of *ZFC does*: see Hamkins (2012a) and Gitman and Hamkins (2011) for details. These subtleties are unimportant for the discussion here, no matter what the axioms taken to characterise the Multiverse are, the arguments carry over immediately.

concepts, identified with models, *via* description results in a very tight link between *understanding* and *reference to ontology*. Regarding categoricity theorems, he says the following:

> The point is that a second-order categoricity argument, even just for the natural numbers, requires one to operate in a context with a background concept of set. And so although it may seem that saying" 1, 2, 3, . . . and so on, "has to do only with a highly absolute concept of finite number, the fact that the structure of the finite numbers is uniquely determined depends on our much murkier understanding of which subsets of the natural numbers exist (Hamkins 2012a, p. 428)

Here, we see Hamkins argue that indeterminacy in the object language transfers to the metalanguage: if we are unable to pin down objects using our first-order set theory, and our second-order[15] theories are dependent upon an understanding of the former, then a second-order metalanguage is also indeterminate. Hence, for the Hamkinsian, we are always confined to first-order descriptions.[16]

Could the Hamkinsian hold the metalanguage fixed? While there is no logical contradiction in doing so, we shall assume that she accepts indeterminacy in her metalanguage (and indeed metametalanguage *et cetera*). There are two reasons for this choice. First, we are interested in Multiversism in its most radical form, and so would like to examine the case where the metalanguage is also indeterminate. Second, to hold the metalanguage determinate seems anathema to the Hamkinsian's dialectical position with respect to the Universist. Recall that some of the main motivations for rejecting the Universist position were his acceptance of seemingly ad hoc interpretations of model-theoretic constructions (especially forcing) for creating alternative set-theoretic universes. An insistence on the determinacy of the metalanguage when the object language is taken to be highly non-absolute would thus seem like her own set of ad hoc assumptions. This is especially so given that the standard way of characterising the metalanguage is in set theory itself. Further, and more conclusively, if the Hamkinsian were to baldly assert determinateness of the metalanguage, it seems that her opponent would be perfectly entitled to likewise claim that his understanding of various set-theoretic concepts are determinate (such as the power set operation) opening the door to categoricity arguments and a rejection of Hamkinsian Multiversism.

We can then see that talk of moving from one universe to another cannot be, strictly speaking, what is occurring on a given occasion of set-theoretic reference. For, by several metatheoretic results, first-order theories are completely unable to precisely determine their objects up to isomorphism.[17] Hence, we lack the conceptual resources to pin down a single universe precisely. We do not, therefore, literally 'pick out' a

[15]Or, indeed, any substantial increase in expressive resources in the metalanguage (e.g. ancestral logic).

[16]This is not to say that second-order set theory is *meaningless* for the Hamkinsian. Rather, the interpretation of the second-order variables is dependent upon background concept of set, and so indeterminacy in the first-order language of *ZFC* carries over to the second-order language.

[17]For example, the Löwenheim–Skolem Theorems entail that if a theory has an infinite model, then it has (clearly non-isomorphic) models of every infinite cardinality. The existence of non-well-founded models (an immediate consequence of the Compactness Theorem for first-order theories)

model, perform a construction on said model, and move to a different model: we lack sufficient and determinate conceptual resources to accomplish such a task. One way to understand Hamkins' suggestion is to hold that we refer to several universes at once *via description*.[18] Since second-order resources are indeterminate as to interpretation of the variables, these descriptions are *first-order*, and as such do not specify any one universe precisely. Set-theoretic reference within the multiverse should thus not be thought of as referring to a particular set-theoretic possibility and then moving from this *via* a construction, but rather identifying a 'cloud' of universes within the multiverse all of which satisfy some particular sentences of first-order set theory. We then perform a model-theoretic construction, relative to each model in the cloud, which shifts focus to a different cloud of universes within the multiverse. We can then still hold on to our understanding of the practice of set theory, it is just that 'V' refers to some (contextually fixed) collection of universes, rather than a single one, and construction is *relative* to each background. This then moves us to a different cloud of universes satisfying different sentences.

Similar views receive currency in the philosophical literature, and have been seen to solve philosophical problems. For example Balaguer (1998) sketches a similar philosophical view (his affinity with which Hamkins notes[19]) on which any coherent first-order description correctly describes part of mathematical reality. In particular Balaguer sees his view as solving the following classic philosophical problem with objects-based realist accounts:

> Benacerraf (1973) (Benacerraf's Problem) How can we refer to mathematical objects at all if we cannot come into direct epistemic contact with them?

Of course, the exact exegesis to be given to this problem is a thorny philosophical issue in itself. Benacerraf's original statement relies explicitly on a causal theory of knowledge, an account few would now accept. However, we may take the problem as well posed: given an acceptance of realism and the existence of mathematical objects as mind-independent, aspatio-temporal entities, how do we gain knowledge of such objects?

The question is of particular interest for the Hamkinsian, as it presents a possible advantage over the Universist. While Universists have proposed various solutions to the problem, either by denying it is actually a problem (Maddy 2011), providing a categoricity argument (McGee 1997); (Martin 2001), or positing a Gödelian faculty of intuition (Gödel 1947), all are fraught with controversy. If she can motivate the claim that she has a solution to the above problem, then this would place her in a far stronger dialectical position.

Balaguer sees his view as responding to the following characterisation of Benacerraf's problem:

(Footnote 17 continued)
is another good example of clearly non-isomorphic models of theories that are the same with respect to first-order satisfaction.

[18] It should be noted that he states that this is only 'often' how reference is achieved. However, nowhere else is he explicit about exactly *how* reference occurs.

[19] See, for example, his presentation in Hamkins (2012b).

> in order to salvage their view, mathematical platonists have to explain how human beings can acquire knowledge of abstract mathematical objects, given that they are not capable of coming into any sort of contact with such objects, that is, receiving any information from such objects (Balaguer 1998, p. 48)

To which he responds as follows:

> Thus, in connection with the epistemological problem, the point is this: if I know that some theory truly describes part of the mathematical realm, then I have knowledge of that realm, regardless of whether it describes a unique part of that realm (Balaguer 1998, p. 50)

The similarity with Hamkins' view is clear: we have first-order descriptions that (non-uniquely) identify part of the mathematical realm. Balaguer and Hamkins' positions are thus similar in content whilst motivated by different considerations. Balaguer attacks an epistemological problem concerning knowledge of mathematical objects in light of Benacerraf's Problem, whereas Hamkins is motivated by worries (both epistemological and ontological) arising from the vast proliferation of set-theoretic models. Common to both the ontological interpretation of Hamkinsian Multiversism and Balaguer's view is that knowledge of mathematical objects can be accounted for by understanding first-order reference to apply to part of the mathematical realm.

Comparison of the ontological interpretation of Hamkinsian Multiversism and Universism is thus particularly interesting from a philosophical perspective, because both are paradigms of *realism*. Each seeks to underpin mathematical practice and truth by having sets be abstract objects in ontologically good standing. However, both views have very different conceptions of the nature of the subject matter of set theory, a fact which is manifest in their attitudes to the bivalence of statements such as the Continuum Hypothesis. The Universist should think that CH has a determinate truth value, holding as they do that there is a single determinate maximal universe of sets to which we may precisely refer. However for the Multiversist there is no fact of the matter whether or not CH is true, some set-theoretic backgrounds that satisfy ZFC also satisfy CH, and some do not. In particular any universe V has a forcing extension $V[G]$ (in which no cardinals are collapsed) such that $V[G] \models ZFC + \neg CH$ and a different extension $V[H]$ (adding no new reals) such that $V[H] \models ZFC + CH$. Asking "Is CH true?" is comparable to asking whether the Parallels Postulate is true in geometry: just as the Parallels Postulate depends on with *which* geometry one is concerned, so CH depends on the particular set-theoretic background under consideration.

11.3.2 The Algebraic Interpretation

The ontological intepretation is not, however, the only way of interpreting Hamkins' view. We begin by noting a difference between two kinds of mathematics. On the one hand we have *particular* mathematics such as number theory and analysis, where we (at least *prima facie*) aim to talk about some particular structure. On the other hand, we have *algebraic* mathematics, on which we explicitly are not concerned with any particular objects, rather we examine what will be possible on a given structure with certain properties.

To see the distinction more clearly, consider a paradigm case of particular mathematics such as number theory. Here we aim to talk about properties held by individual numbers and the natural number structure. For example, we can claim that the number 7 is prime, and in so doing make a claim about properties held by a particular position in the structure $\mathbb{N}$. On the other hand, a paradigm example of algebraic mathematics is group theory, where we study the basic properties of any structure satisfying the group axioms (and possibly extensions thereof). We are not interested in what the objects are, rather we concern ourselves with what will hold and can be constructed given that we are presented with a structure with certain properties. Hamkins, however, occasionally argues that on his view set theory is more of a piece with disciplines such as group theory:

> While group theorists study groups, ring theorists study rings and topologists study topological spaces, set theorists study the models of set theory (Hamkins 2012a, p. 418)

How should we understand the practice of algebraic mathematics? It is tempting to understand it in a similar way to the ontological interpretation of Hamkinsian Multiversism: by providing descriptions of what it takes to be a group or a ring we might be said to describe part of the mathematical realm. Certainly, one could take this view of algebraic mathematics, and in doing so would collapse the algebraic interpretation of Hamkinsian Multiversism into the ontological interpretation. However, especially as the end argument will be that the ontological interpretation is philosophically incoherent, we shall see if there is an alternative understanding available.

A different way to interpret algebraic mathematics such as group theory views the disciplines not as concerned with description and reference, but rather axiomatising facts about what holds and can be constructed relative to certain kinds of mathematical structure. In the case of group theory, if we take a group G, we are able to investigate (a) what holds on G, and (b) what we can construct from G. For example, we know that if G has a prime number of elements, then G is cyclic. Further, we can construct new objects from G. Taking quotients, direct products, semidirect products, wreath products, direct limits, inverse limits *et cetera* are most naturally understood as methods of constructing new groups from old. We can understand these operations not as making any claims about existence and reference, but rather telling us what will happen given some objects endowed with some operations and relations.

The idea then for the Hamkinsian is to hold that her view is not concerned with existence and reference but rather explaining what will hold and can be constructed given some structure that satisfies ZFC. We do not make any claims as to what exists within the Multiverse, rather it is seen as an intuitive picture to facilitate algebraic reasoning concerning sets. Given a structure, the Multiverse View tells us how we can move from this structure. Indeed it is natural to view some of the technical achievements of the Multiversist programme in this light. For example, we can understand the project of providing the modal logic of forcing in (Hamkins and

Loewe 2008) as an attempt to explain how we can move from a ground model using forcing constructions.

It should be noted that *prima facie* the two interpretations are *not* in conflict with one another. Indeed, it is very natural to adjoin the algebraic interpretation to the ontological interpretation in providing an account of exactly what the Hamkinsian is advocating concerning set-theoretic practice. Despite this, the rest of the paper will argue that the ontological interpretation is unsatisfactory (without a softening of the position). Nonetheless, we shall see that the algebraic understanding of the Hamkinsian is not thereby threatened.

11.4 Relativism and the Referential Regress

Let us return then to the ontological interpretation of Hamkinsian Multiversism. Recall that on this interpretation we refer *via* description to a cloud of universes within the Multiverse.

We should first note that Hamkins is committed to a very strong form of *relativism*: what sets exist and the kinds of model-theoretic constructions possible are relative to a particular initial set-theoretic background. Given this, the extent of the multiverse is also indeterminate until a particular point is chosen. This is a fact of which Hamkins is well aware:

> although it is better to understand these descriptions [forcing *et cetera*] as relative construction methods, since the resulting universe described depends on the initial universe in which the constructions are undertaken...one does not expect the properties of the multiverse to be available when undertaking an internal construction within a universe. That is, we do not expect to see the whole multiverse from within any particular universe (Hamkins 2012a, p. 417)

That is to say, the way the Multiverse itself appears differs according to the universe in which one starts. One immediate response to the Hamkinsian would then be the following argument. We might note that, as the extent of the multiverse is determined by set-theoretic background, the following statement is true:

> We cannot quantify over the whole multiverse

Now, this particular statement is clearly true for Hamkins. Let 'extent$_{V_1}$ of the multiverse' denote the extent of the multiverse from the point of view of V_1. We know that for any extent$_{V_1}$ of the multiverse, there is a different extent$_{V_2}$ of the multiverse that contains universes not in the extent$_{V_1}$ of the multiverse. Thus there are always universes in 'the whole multiverse' over which we cannot quantify given the fixing of some background concept of set V. The statement is, at first blush, somewhat paradoxical: it violates its own content by quantifying over the whole multiverse.

We should note, however, that such a paradox is merely intuitive and informal. Since there is no known system in which the notion of the whole Multiverse is formalisable, claims about the Multiverse are formulated in terms of set models within some initial starting background V. Thus, with respect to formal analysis, there is no direct contradiction. For, *within* some V we can perfectly well quantify over 'the whole Multiverse': V simply takes the Multiverse to be some models satisfying the Multiverse Axioms. Given the lack of direct contradiction, we might question the extent to which the Hamkinsian Multiversist will find such arguments convincing. Further, such relativistic paradoxes are well worn in the literature on absolute generality. I will not, therefore, re-examine those issues here. However, I will press a different worry on exactly how the Multiversist is able to secure reference to any particular universe within an extent$_V$ of the multiverse.

We begin by observing that the model-theoretic constructions possible are dependent upon the particular background of set in which one finds oneself, let it be denoted by 'V_1'. Thus V_1 determines the extent$_{V_1}$ of the multiverse. Given this, the exact cloud c_1 picked out by a given utterance of set-theoretic statements is in fact dependent on V_1: if the extent of the multiverse had been different, we would have picked out a different cloud. But as it was acknowledged earlier, picking out an exact universe (V_1) is impossible as we are restricted to referring to models using concepts expressed as first-order axiomatisations. Thus it turns out that the initial selection of c_1 was dependent on a prior selection of a different cloud c_2. Again though, selection of c_2 is dependent on the extent of the multiverse, and hence on selection of a different universe V_2 to determine the extent$_{V_2}$ of the multiverse. But reference to V_2 is impossible, and hence this should really be analysed as reference to some cloud c_3. It is clear that there is no end to this process, and we have a non-well-founded dependency chain.

Clearly a referential regress looms. Our reference depends upon an infinite descending sequence of cloud and universe specifications. Normally (though not always), in philosophical discussion, the production of a regress in a situation where one would expect grounding (such as with respect to reference) is sufficient to show that the view faces some serious difficulties. However, it is one thing to show that a regress looms, and another to show that it is vicious in nature. This is particularly so for the case at hand as, by and large, the Multiversist is happy with the idea of non-well-foundedness. Indeed, a central tenet of the Multiverse View states that every universe is non-well-founded from the perspective of a different universe.[20] So, is the regress vicious?

One might think that it is not, in fact, a vicious regress. While one might worry that in the case of reference a non-well-founded dependency structure is unacceptable, there are examples that show that this is not implausible. Imagine, for example, a non-actual, non-atomic, possible world composed entirely of ever smaller particles of increasing fundamentality. Let 'o_1' denote some object in this world. One might think that we successfully refer to o_1 in virtue of successfully referring to all its smaller parts of the next level of greater fundamentality o_2', o_2'', o_2''', ... We might then think that we refer to these in virtue of referring to all *their* parts of the next level of fundamentality $o_3', o_3'', o_3''', \ldots$ and so on. Here, the non-well-founded dependency structure of reference is not obviously vicious.

[20]This is known as the Well-foundedness Mirage. See Hamkins (2012a), p. 439.

However, epistemological considerations tell us that in the case of the Multiversist it is indeed a vicious regress. Recall the Multiversist's position regarding the link between *ontology* and *concepts*. There it was noted that the Multiversist thought that every model of set theory constituted a set *concept*. Now, it was then noted that, by the Multiversist's own lights, this could not quite be the correct story. The impossibility of singling out individual set-theoretic backgrounds meant that it was far more accurate to say that a set concept is related to a particular cloud of universes within the multiverse. However, now it has been shown that such an occasion of reference requires picking out infinitely many such clouds. The problem is now very serious: given that there is a regress it seems that we are employing *infinitely many concepts* on a given occasion of reference. The idea that we could possess infinitely many concepts is highly controversial (if not downright absurd), let alone *employ* infinitely many concepts in referring. The situation in the non-atomic world is substantially *disanalagous*: there we were only using *one* concept to refer (the concept that sufficed to pick out o_1), it just happened to be the case that this reference had a non-well-founded dependency chain.

Is there a way for the Multiversist to modify their view such that they can avoid this problem? I see two main ways of achieving this goal:

1. Modify her view so that the regress is no longer vicious.
2. Modify her view in such a way as to prevent the generation of the regress at all.

What are the prospects for the first option? If we analyse the non-atomic universe, we can see that the need for infinitely many concepts avoided in virtue of the fact that the speaker at the world did not need to possess the concepts for the more fundamental objects in order to secure reference: in this case reference to o_1 was achieved *via* direct contact with it and all its parts of greater fundamentality. The need for infinitely many concepts is precisely what is required for the Multiversist, in order to know about which cloud we are talking, we must first know what other clouds we are fixing in order to secure reference to the initial cloud, and every such cloud is described by a concept. The Hamkinsian could then possibly avoid the viciousness of the regress by denying the link between concepts and ontology: they could deny that every cloud in the chain is described by a concept.

This response is deeply problematic for familiar reasons in the philosophy of mathematics. In the case of the non-atomic universe, the speaker has direct contact with objects of the kind to which she is referring. This is precisely not so with the proposed account of reference in higher set theory: the way we refer to these entities is entirely through description. If we deny that we have a concept for referring to each cloud, there would be a point at which we did not have a fixed background, and hence reference would break down. Denying the link between concepts and ontology undercuts the Multiversist's initial starting point for reference.

The second option then looks slightly more promising: we should modify the Multiverse View to prevent the regress getting going at all. Recall that the catalyst for the regress was the fact that the *extent* of the Multiverse itself was relative, this then infected our reference to any particular cloud. Two responses present themselves:

either we can argue that this sequence of reference nonetheless has a stopping point despite the indeterminacy in the extent of the Multiverse, or alternatively prevent the regress by delimiting absolutely the extent of the Multiverse.

How could we accomplish the former? Instead of maintaining that the extent of the Multiverse is dependent on where we start, she could instead say that its extent is indeterminate in virtue of being so dependent. On a given occasion of set-theoretic reference, while there may be a potential regress, we simply do stop somewhere, it is just that we do not know where we have stopped. The Multiverse is then fixed from this particular stopping point, which may be different on any particular occasion.

Certainly, there is no contradiction in asserting this. There is nothing that she has said that prevents her from maintaining that we simply end up referring to a particular universe to act as our reference frame for talking about the Multiverse. Such a response is, nonetheless, highly dialectically ineffective. We should be mindful here of the Hamkinsian's complaint against the Universist that he is ad hoc in his disavowal of the existence of legitimate universes other than V. For, the Hamkinsian can give no particular reason to focus on one stopping point rather than another (nor any determinately understood language in which we could try to differentiate stopping points). This process of stopping at some V would thus be to admit her own ad hoc assumptions. Further, such a response would undermine the purported advantage the Multiversist wishes to establish over the Universist concerning Benacerraf's Problem. Salient here is the fact that the Hamkinsian sees herself as providing a response to this central issue by accounting for reference through description without positing the existence of non-natural mental powers (as often Gödel is accused of doing). However, the response that we simply stop somewhere (without being able to give any reason for a particular stopping point) seems, like Gödel, to ascribe unexplained powers to the human mind.

Thus a preferable option is to be able to genuinely single out particular set-theoretic backgrounds as more privileged than others, thereby securing the non-relativity of the model-theoretic constructions. In order to do this, we must have some model or models privileged within the multiverse, such that the key model-theoretic constructions are absolute with respect to these models. In turn, to facilitate this, we require a stock of absolutely understood concepts, sufficiently rich in character that we can identify determinately a class of set-theoretic backgrounds.

The key point to take away from the discussion is the following: in any philosophy of mathematics that tries to make sense of a multiverse picture of set theory as concerned with reference, there is likely to be an element of relativism of certain concepts. However, there is a limit to how far this relativism can go. In particular we need to be precise about which universes (off the bat) are to count as privileged, and in doing so will have to specify a list of concepts that we are simply taking to have absolute significance in order to restrict our class of models. Moreover, this stock of concepts must be sufficiently rich to allow the relevant model-theoretic techniques to be absolutely specified, and the multiverse given precise limit. The acceptance of 'any particular' first-order model as a legitimate universe results in a relativism so strong that it cripples our ability to refer at all.

11.5 Relativism and Metalogic

Consequences of this relativism can be pushed further to create more problems for the Multiversist. Due to the centrality of the claim that no-one universe of sets is more ontologically privileged than another for the Hamkinsian, several key concepts that many would take to be determinate turn out to not be so. Two good examples are the following[21]:

> **Well-foundedness Mirage.** Every universe V is non-well-founded from the perspective of another universe W.
>
> **Countability Principle.** Every universe V is countable from the perspective of another universe W.

Why do the above principles hold? The Well-foundedness Mirage follows from several metatheoretic results: we know that there are models of *ZFC* on which the 'membership relation' is non-well-founded from the perspective of the 'standard' model.[22] Many of these models will be able to see (from the perspective of their 'membership relation') a descending 'membership' sequence in what we currently take to be the 'standard' model, and hence view the current model as non-well-founded.[23] In turn the Countability Principle holds because *any* cardinal can be collapsed to $\aleph_0$ using forcing arguments.

The ramifications for the Multiverse View are both striking and immediate. Any properties that are not absolute between all models of *ZFC* turn out to be inherently dependent on our background concept of set. In this sense, the Multiversist's relativism about set-theoretic background extends to a relativism about many concepts ordinarily taken to be well understood. This is certainly controversial, but a feature of the view Hamkins is happy to embrace: he feels it makes sense of our experience with the diverse the set-theoretic possibilities provided by the model-theoretic constructions. However, we should be mindful of exactly how far this relativism goes. In particular, the relativism even applies to our understanding of natural numbers. Hamkins says the following:

> So why are mathematicians so confident that there is an absolute concept of finite natural number, independent of any set-theoretic concerns, when all of our categoricity arguments are explicitly set-theoretic and require one to commit to a background concept of set? My long-term expectation is that technical developments will eventually arise that provide a forcing analogue for arithmetic, allowing us to modify diverse models of arithmetic in a fundamental and flexible way, just as we now modify models of set theory by forcing, and this development will challenge our confidence in the uniqueness of the natural number structure, just as set-theoretic forcing has challenged our confidence in a unique absolute set-theoretic universe (Hamkins 2012a, p. 428)

---

[21] Hamkins (2012a), pp. 438–439.

[22] Examples of this sort include the Mostowski Collapse Lemma, Compactness Theorem for first-order theories, and Ultrapower Construction (the former implies the existence of non-well-founded models indirectly, the latter two by explicit construction).

[23] The astute reader will notice that the explanation given here is riddled with 'scare quotes'. There is good reason for this, even the notion of standardness itself is relative for the Hamkinsian Multiversist.

Here we see Hamkins express doubts that there is any determinate concept of *natural number*, on the grounds that any attempt to determine the structure of the natural numbers up to isomorphism will have to use resources beyond first-order, and hence will be dependent on the background notion of set. This lack of confidence in an absolute concept of natural number, though it brings out another controversial aspect of the view, is fine as far as it goes. However, the thesis that the natural numbers are a relative concept points to a familiar objection to views that hold that purely first-order theories are the only legitimate expressive resources. We begin with the following well-known theorem:

Theorem 1 Let S be a set of first-order sentences and let $\phi(x)$ be any first-order formula containing only x free. If, for every natural number n, there is a model of S in which the extension of ϕ contains at least n-many objects, then there is a model of S in which the extension of ϕ contains infinitely-many objects.

The theorem (a quick consequence of the Compactness Theorem) has the following immediate consequence: no purely first-order theory can ever pin down the notion of finiteness up to isomorphism.[24] Now real problems are beginning to emerge for the Multiversist. For, the formal framework of her view is to study the Multiverse through analysing the models of *ZFC*. Hence, she would like there to be a well-understood and *determinate* notion of *proof* and *well-formed formula*. Without these, the notions of a proof in *ZFC* and indeed *ZFC* itself are not even determinate. It is therefore indeterminate what the Multiversist is even asserting in saying that universes are models of *ZFC*, and her formal framework breaks down.

Again, the Hamkinsian could respond by arguing that there is a determinate notion of natural number and finitude *relative* to an initial starting V. Every V thinks that it has the standard natural numbers and well-formed formulae. The relevant model-theoretic constructions thus go through relative to that background. Later, we shall see that the intuition behind this thought helps to clarify the algebraic interpretation. For the moment, however, this response falls flat on the ontological interpretation. There we are concerned with attempting to describe and refer to part of the mathematical realm. Asserting "*ZFC*" gets us nowhere if *ZFC* itself is indeterminate: it fails to tell us anything about the way mathematical reality is as it is not clear what is being asserted.

Again, we might take this as indicative of the fact that certain notions need to be taken to be absolute. It has long been noted that certain mathematical concepts are necessary for the expression of metalogical definitions.[25] By adhering to a very strong form of relativism, the Multiversist undercuts the very concepts required to properly express her own view.

---

[24] Many of these considerations, and in particular the above theorem, are discussed in Shapiro's seminal (Shapiro 1991). For a discussion of the similarities and differences between indeterminacy in the notion of finiteness and the explicitly set-theoretic case, the reader is directed to Field (2003).

[25] Certainly, at least since, Shapiro (1991).

11.6 Giving up on the Benacerrafian Challenge: A Response for the Hamkinsian Multiversist Through the Algebraic Interpretation

We have seen two problems for the ontological interpretation of the Hamkinsian Multiversist, one concerning a referential regress and a second concerning difficulties with metalogic. A solution was outlined: soften the radical nature of the view and take some notions to be determinately understood. I want to now consider a different approach: reject the ontological interpretation and simply continue with the algebraic interpretation. This will, however, force her to give up her view as a response to Benacerraf's Problem.

The solution in question involves rejecting that Hamkins' view should be construed as concerned with ontology and reference, and rather view set theory as an *algebraic* enterprise. Recall the options presented for avoiding the referential regress:

1. Modify the view so that the regress is no longer vicious.
2. Modify the view in such a way as to prevent the generation of the regress at all.

We can take option (2), but rather than trying to maintain an account of reference, simply give up on set theory as the kind of enterprise in which we try to refer. Recall our paradigmatically algebraic discipline: group theory. As noted earlier, when reasoning with group theory we do not take ourselves to be referring to any particular group, rather we are simplying saying what will hold and be possible on any given structure that satisfies the group axioms.

We can view Hamkins' project in a similar light. Rather than taking him to be solving a problem of *reference* through description, we can take him to be telling us what will be satisfied and possible on any structure that satisfies the *ZFC* axioms. We make no claim as to the universes that exist, and indeed cannot speak about the extent of or situate ourselves within the Multiverse. We can, however, say that given some structure V that satisfies *ZFC* certain things will hold and can be constructed relative to V. It is then simply a bad question to ask which universe we are in or what universes there are, much as asking which *exact objects* one is talking about in group theory is misguided.

How does this response solve the earlier problems? With respect to metalogic, it is true that on the algebraic conception we do not have a determinate understanding of natural number, finiteness, well-formed formula, or proof, independent of set-theoretic background. For, while any background V in which we find ourselves provably holds itself to be standard (with *the* standard natural numbers and *the* standard well-formed formulae and proofs) we can easily construct a universe that views V as non-standard. The important point from the perspective of the working set theorist is that any proofs or well-formed formulae they talk about will have the normal properties relative to any particular universe of sets in which they may find themselves, and hence the normal constructions will be possible from that background. In this sense, the attempted response on behalf of the Hamkinsian in the previous section was on the right track. However, by holding that her view makes claims about an ontology to be described, the Hamkinsian Multiversist found herself in trouble, she then needed *ZFC* to have determinate content in order to say something meaningful about the mathematical reality. Here, since our claims are already explicitly relative, there is no such problem. We are not making any claim about mathematical reality, rather we simply state what can be done from within a particular structure.

Moreover, the issue of reference is dealt with very trivially. Under the algebraic interpretation, we do not take set theory as the sort of enterprise that is concerned with referring to objects. Rather it is seen as providing an intuitive framework that underlies an algebraic method of thinking. This then allows us to understand what will be possible on a given structure with certain properties. Thus the problem of reference dissolves: we are not even making the appropriate kinds of claims to be assessed for reference.

However, we should note that Balaguer's response to Benacerraf's Problem is utterly abandoned by this characterisation of Hamkinsian Multiversism. First, on this account of the enterprise of set theory the response seems utterly unnecessary, it is the wrong *kind* of question to be asking about set-theoretic practice. Set theory here is rather viewed as an algebraic discipline not concerned with being 'about' any objects. Second, the characterisation also clearly undercuts the response in virtue of the acceptance of a thoroughgoing relativism with respect to metalogic. This makes it impossible to talk about 'reference *via* description' independent of assuming that one is already situated in some V, from the outside perspective it is not determinate what would even constitute a 'description' given the indeterminateness of formulae.

11.7 Conclusion

We have seen that Hamkins' Multiverse View appears to present a fascinating and elegant way of providing a background ontology for model-theoretic constructions. There appear to be two different aspects to the view, one *ontological* and one *algebraic*. The ontological interpretation of the view is seen to commit the Multiversist to a referential regress, vicious in virtue of the correlation of concepts with models. The problem is instructive for modern philosophical practice and shows that if one wishes to have an account of reference in set theory, a stock of notions taken to be absolute is required in order to secure a preferential class of models and delimit the multiverse. This is further seen with respect to metalogical definitions, as their satisfactory statement requires certain notions (such as finitude) to be absolutely understood. It is possible to maintain the Hamkinsian Multiverse perspective by viewing set theory as a fundamentally algebraic enterprise, but this is at the cost of a response to Benacerraf's Problem and a view of set theory on which we are concerned with *ontology*. While there are options open for the Hamkinsian, the moral is clear: if one wishes to have reference to mathematical objects, then relativism has to stop somewhere.

Acknowledgments The author is very grateful to Ben Fairbairn, Joel Hamkins, Peter Koellner, Ian Rumfitt, and audiences in London and Milan for insightful and useful feedback on the issues discussed. Special mention must be made of Victoria Gitman, Alex Kocurek, Chris Scambler, and two anonymous referees whose detailed comments improved the paper immensely. The author also wishes to thank the Arts and Humanities Research Council for their support during the preparation of the paper.

References

Balaguer, M. (1998). *Platonism and anti-platonism in mathematics*. Oxford: Oxford University Press.

Benacerraf, P. (1973). Mathematical truth. *The Journal of Philosophy*, *70*(19), 661–679.

Feferman, S. (2011). *Is CH a definite problem?* Retrived April 13, 2014, from http://math.stanford.edu/feferman/papers.html

Field, H. (2003). Do we have a determinate conception of finiteness and natural number? In M. Schirn (Ed.), *The philosophy of mathematics today* (pp. 99–130). Oxford: Clarendon Press.

Foreman, M. (2010). Ideals and generic elementary embeddings. In A. Kanamori & M. Foreman (Eds.), *Handbook of set theory* (pp. 885–1147). New York: Springer.

Gitman, V., & Hamkins, J. D. (2011). A natural model of the multiverse axioms. Preprint. arXiv:1104.4450 [math.LO].

Gödel, K. (1947). What is Cantor's continuum problem? In K. Gödel, *Collected works* (Vol. II, pp. 176–187). Oxford: Oxford University Press.

Hamkins, J. D. (2012a). *Pluralism in set theory: Does every mathematical statement have a definite truth value?* Held at *Philosophy Colloquium: Cuny Graduate Center*.

Hamkins, J. D. (2012b). The set-theoretic multiverse. *The Review of Symbolic Logic*, *5*(3), 416–449.

Hamkins, J. D., & Loewe, B. (2008). The modal logic of forcing. *Transactions of the American Mathematical Society*, *360*(4), 1793–1817.

Hamkins, J. D., & Seabold, D. E. (2012). Well-founded Boolean ultrapowers as large cardinal embeddings. arXiv:1206.6075 [math.LO].

Isaacson, D. (2011). The reality of mathematics and the case of set theory. In Z. Noviak & A. Simonyi (Eds.), *Truth, reference, and realism* (pp. 1–76). Hungary: Central European University Press.

Jech, T. (2002). *Set theory*. New York: Springer.

Koellner, P. (2013). Hamkins on the multiverse. Held at *Exploring the frontiers of incompleteness*.

Kunen, K. (1980). *Set theory: An introduction to independence proofs*. Amsterdam: Elsevier.

Maddy, P. (1988). Believing the axioms. I. *The Journal of Symbolic Logic*, *53*(2), 481–511.

Maddy, P. (2011). *Defending the axioms*. Oxford: Oxford University Press.

Martin, D. (2001). Multiple universes of sets and indeterminate truth values. *Topoi*, *20*(1), 5–16.

McGee, V. (1997). How we learn mathematical language. *The Philosophical Review*, *106*(1), 35–68.

Meadows, T. (2013). What can a categoricity theorem tell us? *The Review of Symbolic Logic*, *6*, 524–544.

Potter, M. (2004). *Set theory and its philosophy: A critical introduction*. Oxford: Oxford University Press.

Shapiro, S. (1991). *Foundations without foundationalism: A case for second-order logic*. Oxford: Oxford University Press.

Steel, J. (2014). Gödel's program. In J. Kennedy (Ed.), *Interpreting Gödel* (pp. 153–179). Cambridge: Cambridge University Press.

Zermelo, E. (1930). On boundary numbers and domains of sets. In W. B. Ewald (Eds.), *From Kant to Hilbert: A source book in mathematics* (Vol. 2, pp. 1208–1233). Oxford: Oxford University Press.

Chapter 12
Forcing, Multiverse and Realism

Giorgio Venturi

[I]n the long run, as often happens, philosophical importance and future mathematics go hand in hand.

G. Kreisel

Abstract In this article we analyze the method of forcing from a more philosophical perspective. After a brief presentation of this technique we outline some of its philosophical imports in connection with realism. We shall discuss some philosophical reactions to the invention of forcing, concentrating on Mostowski's proposal of sharpening the notion of generic set. Then we will provide an overview of the notions of multiverse and the related philosophical debate on the foundations of set theory. In conclusion, we connect this modern debate and Mostowski's proposal, suggesting a way to analyze the notion of genericity within the framework of a multiverse structure.

Keywords Set theory · Realism · Genericity

12.1 Introduction

Without a doubt we can consider 1963 as a turning point in the history of set theory. Indeed, it was in this year that Paul Cohen announced the independence of the Continuum Hypothesis, CH (Cohen 1963). The proof of this remarkable fact rested on a completely new technique, the method of forcing, which in subsequent years proved itself to be a versatile tool for obtaining many other independence results. Since then, the plethora of relative-consistency results obtained by forcing, together

G. Venturi (✉)
Centro de Lógica, Epistemologia e História de la Ciência (CLE), Campinas, Brazil
e-mail: gio.venturi@gmail.com

G. Venturi
Institut d'histoire et de philosophie des sciences et des techniques (IHPST), Paris, France

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_12

with the improvements and sophistications of the method, had the effect of making the study of independent phenomena one of the prominent fields of research within set theory.

The strength of forcing consists in producing different models of ZFC in which it is possible to verify or falsify propositions that are therefore shown to be independent from ZFC. From a philosophical perspective, the increase of well-studied conflicting models of ZFC had the consequence of putting into question the absoluteness of the notion of truth in set theory. As an example of this attitude we may consider the opening sentences of Väänänen (2014).

> After the incompleteness theorems of Gödel, and especially after Cohen proved the independence of the Continuum Hypothesis from the ZFC axioms, the idea offered itself that there are absolutely undecidable propositions in mathematics, propositions that cannot be solved at all, by any means. If that were the case, one could throw doubt on the idea that mathematical propositions have a determined truth-value and that there is a unique well-determined reality of mathematical objects where such propositions are true or false.[1]

Without addressing directly the problem of truth in set theory, we would like to attempt a philosophical analysis of the dramatic change of perspective that the invention of forcing instigated in set theory. In particular, we will concentrate on the contemporary formulation of this philosophical debate: the proposal of different multiverse conceptions.

The paper is organized as follows. In Sect. 12.2 we give a general presentation of the method of forcing and outline the most philosophically critical aspects: genericity and the connected realist attitude. In Sect. 12.3 we briefly recall a philosophical debate on forcing that took place in 1965 at the International Colloquium of the Philosophy of Science held at Bedford College, in London. We will place an emphasis on Mostowski's suggestion to sharpen our intuitions about the notion of genericity. In Sect. 12.4 we present the modern multiverse and universe conceptions, analyzing their significance for the relevant philosophical themes. In Sect. 12.5 we suggest how a multiverse can represent a tool for a finer analysis of the notion of genericity.

12.2 The Method of Forcing

The method of forcing can be presented in at least three alternative ways: using countable transitive models (as presented in Kunen 1980); with boolean-valued structures (Bell 2005); or in a finitistic way (see pp. 233–234 of Kunen 1980). Although from a purely mathematical point of view the three methods are equivalent for what concerns the independence results they give rise to,[2] from a more theoretical perspective they differ both in spirit and with respect to the background assumptions.

---

[1] Väänänen (2014), p. 180.

[2] However, because of the presence of absoluteness results between models of set theory, as for example Shoenfield Absoluteness Theorem, forcing may also be considered as a tool to prove theorems in ZFC. As an example, kindly offered by an an anonymous referee, it is possible to show that every countable partial order embeds into the Turing degrees, since in the forcing extension to

For the first method we normally assume the existence of a countable transitive model of ZFC.[3] This assumption, although weaker than assuming the existence of an inaccessible cardinal,[4] is not only stronger than the sentence expressing the consistency of ZFC—which we shall write Con(ZFC)—but it also implies Con(ZFC + Con(ZFC)), Con(ZFC + Con(ZFC + Con(ZFC))), and so on transfinitely.[5]

For what concerns the boolean-valued models approach, although it allows to internalize the forcing technology inside ZFC,[6] nonetheless this presentation relies on the existence of an evaluation function that cannot be entirely formalized in ZFC, due to Tarski's theorem on the undefinability of truth. More concretely, given a complete boolean algebra $\mathbb{B}$, a boolean-valued structure is defined by transfinite recursion as

$$V^{\mathbb{B}} = \{x \in V : \exists \alpha \in Ord(x \in V_{\alpha}^{\mathbb{B}})\}$$

where, letting α be an ordinal, $Fun(x)$ an abbreviation for "x is a function", $Ran(x)$ and $Dom(x)$, respectively, the range and the domain of x,

$$V_{\alpha}^{\mathbb{B}} = \{x \in V : Fun(x) \wedge Ran(x) \subseteq \mathbb{B} \wedge \exists \xi < \alpha(Dom(x) \subseteq V_{\xi}^{\mathbb{B}})\}.$$

Notice that $V^{\mathbb{B}}$, technically specking, is not a model, being instead a proper class.

Extending the language of set theory with constants for every element of $V^{\mathbb{B}}$, we can define a new language $\mathcal{L}^{\mathbb{B}}$ that allows us to talk about the boolean valued structure $V^{\mathbb{B}}$ by means of linguistic expressions that are then called $\mathbb{B}$-sentences. The main motivation of the above definition consists in allowing a generalization of the notion of satisfaction. Indeed, it is possible to define a $\mathbb{B}$-evaluation function that assigns to every $\mathbb{B}$-sentence σ, its Boolean truth value[7] $[\![\sigma]\!]^{\mathbb{B}} \in \mathbb{B}$, and that makes it possible for the validity of all axioms of ZFC to hold, meaning that if φ is one of them, then $[\![\varphi]\!]^{\mathbb{B}} = 1_{\mathbb{B}}$ i.e.; the maximal element of the complete boolean algebra

(Footnote 2 continued)
add mutually generic Cohen reals we may find a universal order where they embed and since the existence of such structure is Σ^1_2, by the Shoenfield Absoluteness Theorem, its existence can be inferred in the ground model. In the context of strong set theoretical hypotheses we may interpret some of the results from Friedman (2000) and Viale (2016) as using forcing as a tool to prove theorems.

[3]Notice that it may also possible to work in ZFC and, by means of Levy's reflection theorem, to use countable transitive models of finite fragments of it.

[4]See Jech (2002) for the undefined set theoretical terms we will use. However, we will try to make this presentation as self-contained as possible.

[5]See Hamkins (2012), for a discussion on the consistency strength of the assumption of a countable transitive model of ZFC.

[6]Indeed, we may start with M modeling ZFC, and then build $(V^B)^M$, which is a set, and then perform the quotient $(V^B)^M/U$ by an ultrafilter to get an actual model satisfying the forced theory. See Hamkins and Seabold (2012) for details.

[7]Remember that $0_{\mathbb{B}}$ and $1_{\mathbb{B}}$, respectively the smallest and largest elements of a boolean algebra $\mathbb{B}$, may be interpreted as, respectively, falsehood and truth; hence all other elements of the algebra in between have the intuitive meaning of intermediate truth values.

$\mathbb{B}$ that indicates the value "true".[8] The definition of such a $\mathbb{B}$-evaluation function not only resembles closely that of a truth definition, but, similarly, it cannot be fully defined within ZFC, because, as in the case of the truth predicate, the collection of all ordered pairs $(\sigma, [\![\sigma]\!]^{\mathbb{B}})$ is not a definable class in ZFC.[9]

According to the third presentation, if we work in a finitistic setting we do not need to go beyond ZFC, since ZFC proves the existence of models for every finite portion of its axioms. However, either we assume a purely formalistic point of view, or it is hard to argue in favor of a theoretical value of such independence proofs with respect to a notion of truth, *à la* Tarski, intended as a correspondence between syntax and semantics. Moreover, as suggested in Hamkins (2012), this kind of presentation rests on a meta-theoretical definition of finiteness.

Since our aim is to propose philosophical considerations about the application of forcing to mathematics, we believe it is better to concentrate on the presentation of the method via countable transitive models. There are many reasons for this choice, some of which are historical—it is more faithful to the original presentation of the method by Cohen—practical—it is closer to mathematical practice—and theoretical. We already hinted at the theoretical limits of using models of finite fragments of ZFC. On the other hand, although the new countable transitive models produced by forcing are non-standard, the boolean-valued structures are non-standard in a possibly stronger sense. This is because the interpretations of the basic notions of equality and membership have an *ad hoc* non-standard character. Given a complete boolean algebra $\mathbb{B}$ and $u, v \in V^{\mathbb{B}}$,

$$[\![u \in v]\!]^{\mathbb{B}} = \bigvee_{y \in Dom(v)} \big(v(y) \wedge [\![u = y]\!]^{\mathbb{B}}\big);$$

and

$$[\![u = v]\!]^{\mathbb{B}} = \bigwedge_{x \in Dom(u)} \big(u(x) \Rightarrow [\![x \in v]\!]^{\mathbb{B}}\big) \wedge \bigwedge_{y \in Dom(v)} \big(v(y) \Rightarrow [\![y \in u]\!]^{\mathbb{B}}\big).$$

[10]Not only are the above evaluations inter-defined by a simultaneous recursion, but they are also defined in terms of the algebraic structure of the boolean algebra.

We now briefly recall the method of forcing in connection with countable transitive models, outlining those aspects that are conceptually more relevant to a philosophical discussion.

---

[8]See Bell (2005) for the precise proofs of this fact.

[9]However, notice that due to the definability in the ground model of the forcing relation we have that, fixing a formula φ or given a φ of a fixed complexity, the collection $\{(\tau, [\![\varphi(\tau)]\!]) | \tau \in V^B\}$ is a class definable in V^B.

[10]In order to make this presentation shorter and more self contained, following Bell (2005), we use here a functional notation, instead of using $\mathbb{B}$-names, that are more standard in modern presentations of a boolean-valued construction.

12.2.1 Generic Extensions

Given a countable transitive model (c.t.m.) M of ZFC, which we call a *ground model*, and a partial order (poset) $\mathbb{P} \in M$, which we call a *notion of forcing* (or just forcing), the method of forcing produces a new countable transitive model $N \supseteq M$, called a *generic extension*, with the same ordinals of M, i.e. $M \cap Ord = N \cap Ord$. In order to give a more precise description of a generic extension N, we need the following definitions.

Definition 12.2.1 Given a poset $(\mathbb{P}, \leqslant)$, if $p \leqslant q$ we say that *p extends q*. Moreover, we say that a subset $D \subseteq \mathbb{P}$ is dense whenever every element of $\mathbb{P}$ has an extension in D: i.e.

$$\forall x \in \mathbb{P} \exists d \in D(d \leqslant x).$$

Finally we say that a subset $G \subseteq \mathbb{P}$ is a filter whenever $p \in G$ and $p \leqslant q$ together imply $q \in G$, and, moreover, when two of its elements always have a common extension (we then say that they are *compatible*): i.e.

$$\forall p, q \in G \exists r \in G(r \leqslant p \wedge r \leqslant q).$$

Definition 12.2.2 Let M be a c.t.m. of ZFC and let $\mathbb{P} \in M$ be a notion of forcing, we say that a filter $G \subseteq \mathbb{P}$ is M-generic if it intersects every dense subset of $\mathbb{P}$ that belongs to M; i.e. $G \cap D \neq \varnothing$, for all $D \subseteq \mathbb{P}$ such that $D \in M$.

The notion of *genericity* is fundamental for a more precise description of the models produced by the method of forcing. Indeed an M-generic filter G is able to fully determine, together with M, a generic extension, denoted by $M[G]$. The structure $M[G]$ is indeed obtained as the closure of M and G under all set theoretical operations admissible in ZFC. A useful parallel can be seen between this and the construction of $\mathbb{C}$ as the algebraic closure of $\mathbb{R}$ and i. Moreover, assuming that a poset $\mathbb{P}$ has the property that below every condition there are two incompatible conditions (i.e. such that they do not have a common extension in $\mathbb{P}$) we have that the M-genericity of a filter $G \subseteq \mathbb{P}$ implies $G \notin M$. In this case we then have that the model $M[G]$ is a proper extension of M.

The main results that Cohen proved with respect to this kind of construction are the following fundamental theorems.

Theorem 12.2.3 (Cohen 1963) *If M is a c.t.m., $\mathbb{P}$ a poset in M and $p \in \mathbb{P}$, then there is an M-generic filter $G \subseteq \mathbb{P}$ such that $p \in G$.*

Theorem 12.2.4 (Cohen 1963) *If M is a c.t.m., $\mathbb{P}$ a non-trivial poset in M and $G \subseteq \mathbb{P}$ a M-generic filter, then there is a model $M[G]$ of ZFC, such that $M \subseteq M[G]$, $G \in M[G] \setminus M$ and $M \cap Ord = M[G] \cap Ord$.*

Theorem 12.2.5 (Cohen 1963) *If M is a c.t.m. and $\mathbb{P}$ a poset in M, then it is possible to define a relation $\Vdash$, called* the forcing relation*, such that for every M-generic filter G and for any formula φ in the language of set theory we have*

$$\exists p \in G(p \Vdash \varphi) \iff M[G] \models \varphi.$$

Moreover, for a fixed formula φ the relation $p \Vdash \varphi$ is definable in M.[11]

The results quoted above show that generic filters exist, that the closure of M under an M-generic filter G is still a model of ZFC and that truth in the generic extensions can be controlled by a relation definable in the ground model.

What is philosophically relevant here is that assuming the existence of a c.t.m. M we can prove on the one hand that there are new sets outside M, while on the other hand that there are new models of set theory to which these sets belong. Moreover, we can control truth inside these new non-standard models by means of the forcing relation $\Vdash$.

The fact that it is possible to find new sets in the universal class V—i.e. new sets that satisfy the formula $\varphi(x) \equiv (x = x)$, which defines V—outside a given countable model is not surprising because of Cantor's theorem on the uncountability of $\mathcal{P}(X)$, for any infinite set X. However, Theorem 12.2.4 tells us something more about V itself: namely the fact that V is not countable. In other words, Theorem 12.2.4 tells us that the indented interpretation of the axioms of ZFC cannot be a countable transitive model.[12] This fact is non-trivial since the ineffability of V—that is the impossibility to characterize it with any formula of set theory because of Levy's reflection theorem—would leave open the possibility that V itself was countable. But if so, granting a foundational theory in which to run the proof of Theorem 12.2.4, it would be possible to prove the existence of a set outside V, against its universality.

Of course, the above argument is superfluous when ZFC is the background theory; however, the point we would like to make is that the method itself has theoretical consequences on set theory. This is the theoretical context in which we believe it is worth evaluating the philosophical relevance of forcing with respect to set theory.

12.2.2 Realism

This tension between the intended interpretation of the axioms of ZFC and the possibility to construct different models of the theory by means of forcing, is in our view a hint of the presence of a realist attitude that pervades modern set theory, together with its extensive use of forcing, both at a theoretical and at a practical level.

By Cohen's theorem we know that given a countable transitive model, we can force over it and add new sets that lie in the universe of all sets V. As a consequence, it seems that the existence of a set is prior to and independent from the notion of "existence in a model" that naturally fits with the semantics of a first-order theory.

---

[11] As in the case of a $\mathbb{B}$-evaluation function the forcing relation $\Vdash$, intended as the collections of all pairs (p, φ) where φ is in the forcing language, is not definable due to the indefinability of the truth predicate. However, as noticed before, for a fixed $\varphi(x)$ with a free variable the set of (p, τ) for which p forces $\varphi(\tau)$ is definable. Hence the definability of $p \Vdash \varphi$, for a fixed φ.

[12] In fact, neither is it a countable model, as is pointed out in Hamkins (2012).

However, the practice of forcing does not usually refer to countable models of set theory. As a matter of fact, normally the arguments run as follows: take a poset $\mathbb{P}$ with some particular property and force with it over V. Then, the implicit assumption of this abuse of notation is that outside the definable universe of all sets there are ideal objects (sets) that can be found and added to it. This position then betrays an even stronger realist attitude, since sets are considered to exist independently of the intended domain of existence: V.

What we are suggesting is that the use of forcing sustains two different forms of realism that can be roughly described as follows. The first one may be called trans-model realism and consists in viewing a generic extension as a real extension of a countable transitive model of set theory—and not just as a *façon de parler* like the reference to the addition of a new symbol suggests in the case of the algebraic closure of a field—able to capture more of the universe V. This position helps in understanding a common way of referring to the presence of an internal and an external point of view in the application of forcing, which is perfectly exemplified by the next quotation taken from Kunen (1980).

> People living in M cannot construct a G which is $\mathbb{P}$-generic over M. They may believe on faith that there exists a being to whom their universe, M, is countable. Such a being will have a generic G and $f_G = \bigcup G$.[13]

Without discussing Kunen's reference to a being, it is important to notice that what is at stake here is a notion of existence—let us say in V—in a domain that, although not completely formalizable, is approximated by countable models. This notion of existence is prior to the one of existence in a model, since it is the existence of an outer G, in V, that allows the construction of a generic extension $M[G]$. And, as hinted before, it is also sustained by the method itself, which allows the existence of new objects in V, outside M. The more relevant theoretical consequence of such a form of realism is the acceptance of V as an actual existent domain.

The second form of realism we want to describe may be called trans-universe realism and it is a higher-order version of the first. Indeed, assuming an internal point of view and shifting the focus from models to universes, this position accepts the possibility of talking about extensions of the universe of set theory. It is perfectly exemplified by the following quotation from Jech (2002).

> The modern approach to forcing is to let the ground model be the universe V, and pretend that V has generic extensions, i.e. postulate the existence of a generic set G, for the given set of forcing conditions. As the properties of the generic extension can be described entirely within the ground model, statements about $V[G]$ can be understood as statements in the ground model using the language of forcing.[14]

It may be argued that this second form of realism is only an abuse-of-notation version of the first; nevertheless we believe that every such abuse hides a theoretical intention, and that in this case this consists in proposing an absolute and platonic

[13] Kunen (1980), p. 193.

[14] Jech (2002), p. 201.

notion of existence that surpasses the idea that a first-order model is the context in which to give a consistent treatment of a formal notion of existence.

As a consequence, there are two different issues that need to be clarified. On the one hand, is this notion of existence, prior with respect to "existence in a model", compatible with the notion of set exemplified by the ZFC axioms? On the other hand, does the stronger notion of existence, prior with respect to "existence in the intended interpretation", allow for a formal treatment able to supports its use?

12.2.2.1 Set Theoretical Practice

For what concerns the second question we believe that the answer is yes, thanks to the fine analysis produced by Hamkins (2012). Without reporting his argument in its precise formulation, we quote the relevant result that shows that what may appear as an abuse of notation (i.e. referring to the following reasoning: "Let $G \subseteq \mathbb{B}$ be V-generic.[15] Argue in $V[G]$ …") can in fact be codified in set theory by formal means.

Theorem 12.2.6 (Hamkins 2012) *If V is the universe of set theory and $\mathbb{B}$ is a complete boolean algebra, then there is in V a definable class model of the theory expressing what it means to be a forcing extension of V. Specifically, in the forcing language with $\in$, constant symbols $\check{x}$ for every element $x \in V$, a predicate symbol $\check{V}$ to represent V as a ground model, and constant symbol $\dot{G}$, the theory asserts:*

1. *The full elementary diagram of V, relativized to the predicate $\check{V}$, using the constant symbols for elements of V.*
2. *The assertion that $\check{V}$ is a transitive proper class in the (new) universe.*
3. *The assertion that $\dot{G}$ is a $\check{V}$-generic ultrafilter on $\check{\mathbb{B}}$.*
4. *The assertion that the (new) universe is $\check{V}[\dot{G}]$, and ZFC holds there.*

Theorem 12.2.6 consists of the syntactic form of the *Naturalness account of forcing* that is justified by the following result, that is the semantic version of Hamkins *Naturalist approach*.

Theorem 12.2.7 (Hamkins 2012) *For any forcing notion $\mathbb{P}$, there is an elementary embedding*

$$V \precsim \bar{V} \subseteq \bar{V}[G]$$

of the universe V into a class model $\bar{V}$ for which there is a $\bar{V}$-generic filter $G \subseteq \bar{\mathbb{P}}$ (where $\bar{\mathbb{P}}$ is the image of $\mathbb{P}$ under the elementary embedding from V to $\bar{V}$). In particular, $\bar{V}[G]$ is a forcing extension of $\bar{V}$, and the entire extension $\bar{V}[G]$, including the embedding of V into $\bar{V}$, are definable classes in V, and $G \in V$.

[15]Similar to the case of a countable transitive model, a filter $G \subseteq \mathbb{B}$ being V-generic means that G intersects all dense subsets of $\mathbb{B}$ in V: i.e. all dense subsets of $\mathbb{B}$.

This approach to forcing is intended to justify from a formal point of view set theoretical practice. In the view expressed in Hamkins (2012), this approach is also intended to justify a multiverse conception of set theory where many models of set theory are considered to exist in a strong platonic sense.

We will come back later to the general notion of multiverse and we will discuss Hamkins' position. For now, we just mention that this form of naturalism can hardly justify an ontological conclusion due to its schematic character, as acknowledged in Hamkins (2012), p 425.: "This is really a theorem scheme, since V does not have full uniform access to its own elementary diagram, by Tarski's theorem on the undefinability of truth. Rather, the theorem identifies a particular definable class model, and then asserts as a scheme that it satisfies all the desired properties". On the contrary, we agree with Hamkins about the epistemological strength of the above result: we can argue as if we were extending the universe V because what really happens—from the point of view of a first order axiomatization of set theory—is that we are extending what we know about V. In this reading the naturalist approach really mirrors practice and maintains that there are real extensions of partial descriptions of V.

12.2.2.2 The Notion of Set

For what concerns the question of whether the notion of set exemplified by the ZFC axioms is compatible with a notion of existence stronger than "existence in a model", let us have a more historical focus, and consider the ideas of Cantor and Zermelo.

In Cantor's work the first definition of set is from 1882, in the third paper of a series of six from the period 1878–1884 bearing the title *Über unendliche, lineare Punktmannichfaltigkeiten*.

> I call a manifold (an aggregate [Inbegriff], a set) of elements, which belong to any conceptual sphere, well-defined, if on the basis of its definition and in consequence of the logical principle of excluded middle, it must be recognized that it is internally determined whether an arbitrary object of this conceptual sphere belongs to the manifold or not, and also, whether two objects in the set, in spite of formal differences in the manner in which they are given, are equal or not. In general the relevant distinctions cannot in practice be made with certainty and exactness by the capabilities or methods presently available. But that is not of any concern. The only concern is the internal determination from which in concrete cases, where it is required, an actual (external) determination is to be developed by means of a perfection of resources.[16]

It is important to notice that in the above definition there is a reference to a "conceptual sphere" that disappears in the subsequent definitions. Indeed, in Cantor's mature work there is no reference to the universe of all sets because it is the context of discourse, outside which nothing can exist. Then we can say that in Cantor's conception the notion of set depends on its domain of existence; local, before, and global, after. Notice that a notion of existence prior to the one of "existence in a

[16] Zermelo (1932), p. 150.

model" is in principle compatible with a global notion of existence in the intended interpretation of the collection of all sets, that, as acknowledged also by Cantor, cannot be subject to a purely formal treatment due to the presence of the paradoxes. Since it is a difficult task to match Cantor's pre-formal notions and the set theoretical notions we have after the axiomatization of the theory, it is more instructive to turn to Zermelo's ideas.

In 1908, Zermelo began his axiomatization of set theory in the context of Hilbert's school. In the first lines of Heijenoort (1967) we read "Set theory is concerned with a *domain* $\mathcal{B}$ of individuals, which we shall call simply *objects* and among which are the *sets*."[17] The difference between Cantor's late definitions and Zermelo's is to be found not only in the absence, in 1908, of a concept-extension notion of set, but also in a local conception of set theory, instead of a global one. The notion of domain would remain constant in Zermelo's reflection, and culminated in 1930 with the work *On boundary numbers and domains of sets* where a quasi-categoricity theorem for ZF will be proved. In this article, where the birth of a modern approach to the universe of sets in terms of a cumulative hierarchy can be found, the difference between sets and domains becomes less marked, because: "every categorically determined domain can also be conceived of as a 'set' in some way or another".[18] Then, together with the consequent change of the underlying notion of set, the difference between sets and domains becomes a matter of context. As a consequence, the possibility that sets may exist outside a domain is granted by the indefinite possibility to extend a domain, turning it into a set.

In conclusion, we can say that the notion of set that underlies the use of forcing in set theory is compatible with the one that inspires the modern formulation of its axiomatization: an iterative conception exemplified by a cumulative hierarchy. In particular, this notion, and hence also the method of forcing—at least in its presentation by means of countable transitive models—sustains a realist component that originates from the use of a notion of existence stronger than the one given by first order logic: i.e. "existence in a model".

There is, however, an aspect that has not been clarified so far: the choice of the logical context of our reflections on the realist components of the forcing method. This is relevant since a second-order logical approach may trivialize the realist position sustained by the existence of generic extensions, by maintaining that the existence of a generic filter is not given by the method itself, but by the fact that we are—or should be—dealing with structures that satisfy ZFC_2, the second-order version of the axioms of ZFC. In other words, the existence of a generic object may not granted by the method, but by our background theory.

As we have already hinted at the end of Sect. 12.2.1, our goal is to give a philosophical account of forcing by making clear which are its theoretical consequences on set theory. As a matter of fact, we interpret a philosophical analysis of the method of forcing as a tool for better understanding the notion of truth in set theory; hence we would like to keep our initial assumptions as sober as possible. Our aim is not

[17] Heijenoort (1967), p. 201.
[18] Zermelo (2010), p. 429.

only to understand the mathematical results, but also the theoretical information we can grasp from the use of forcing. Moreover, there is also an even stronger reason for concentrating on first-order logic in connection with a philosophical study of forcing: the fact that in second-order logic CH is decided, while forcing has been created and used in order to show and understand the limits of first-order logic. In other words, we believe that a second-order look, in this context, may be a methodological mistake.

Finally we would like to make clear that the reference to Zermelo's notion of set, in arguing in favor of the accordance between the trans-model realism of forcing and the modern conception of set, does not negatively affect the choice of first-order logic as the context of our inquiry. Indeed, it is not the second-order character of Zermelo's notion of set, but his cumulative hierarchy picture of the universe of sets, which gave rise to the iterative conception, that suffices in arguing that the notion of existence of a set is stronger than that of "existence in a model".

This way of conceiving mathematical objects has been named *quasi-combinatorial*, by Bernays in 1935 (Bernays 1983).

> But analysis is not content with this modest variety of platonism [to take the collection of all numbers as given]; it reflects it to a stronger degree with respect to the following notions: set of numbers, sequence of numbers, and function. It abstracts from the possibility of giving definitions of sets, sequences, and functions. These notion are used in a 'quasi-combinatorial' sense, by which I mean: in the sense of an analogy of the infinite to the finite.
>
> Consider, for example, the different functions which assign to each member of the finite series $1, 2, \ldots, n$ a number of the same series. There are n^n functions of this sort, and each of them is obtained by n independent determinations. Passing to the infinite case, we imagine functions engendered by an infinity of independent determinations which assign to each integer an integer, and we reason about the totality of these functions.
>
> In the same way, one views a set of integers as the result of infinitely many independent acts deciding for each number whether it should be included or excluded. We add to this the idea of the totality of these sets. Sequences of real numbers and sets of real numbers are envisaged in an analogous manner. From this point of view, constructive definitions of specific functions, sequences, and sets are only ways to pick out an object which exists independently of, and prior to, the construction. The axiom of choice is an immediate application of the quasi-combinatorial concepts in question.[19]

As a matter of fact, the above quotation, in Bernay's intentions, was intended to exemplify the underlying notion of set that was formalized by Zermelo in his axiomatization of set theory. Since we argued that this notion perfectly matches with the theoretical ideas connected to the most philosophically significant presentation of the method of forcing, we believe not only that the central notion of genericity deserves a philosophical inquiry, but also that this analysis may shed lights on the realist component of set theoretical practice.

[19] Bernays (1983), pp. 259–260.

12.3 The Initial Reactions

In 1965, soon after Cohen's results were made public, the International Colloquium in the Philosophy of Science—held in London in July of that year—was one of the first opportunities for a collective philosophical discussion on the meaning of the independence results obtained by Cohen's method and on their effect on set theory. In the section *Problems in the Philosophy of Mathematics* we can find—among the nine contributions given by A. Robinson, A. Mostowski, P. Bernays, S. Körner, G. Kreisel, Á. Szabó, L. Kalmár, J. Easley and F. Sommers—five talks that referred directly to Cohen's theorems. Moreover, three of them dealt directly with the question of the philosophical interpretation of these results: the ones of Mostowski, Bernays and Kreisel. It is possible to find their contributions, together with the subsequent discussions, in Lakatos (1967).

One topic of discussion, sponsored mainly by Kreisel, was the discrepancy between the negative results given by Cohen and the possibility to have a quasi-categorical axiomatization of ZFC in second-order logic, where the Continuum Hypothesis finds a definitive, although unknown, solution. As a consequence, the concerns motivated by the independence results given by Cohen's method were ascribed to the limits of the first-order axiomatization of set theory. On this matter Bernays's paper is a lucid example: "The first essential thing which emerges from his paper is that the results of Paul J. Cohen on the independence of the continuum hypothesis do not directly concern set theory itself, but rather the axiomatization of set theory".[20]

Although from different philosophical positions, many authors involved in that discussion acknowledged that the independence of CH cannot count as a definitive solution to the Continuum Problem. An interesting suggestion of comparison, which will be also considered later in connection with the multiverse discussion, was between the independence phenomena in set theory and in other branches of mathematics, such as arithmetic,[21] group theory or geometry. In particular, CH was contrasted with Euclid's Fifth postulate, and the main difference between the independence of the two propositions was considered at a logical level: Euclid's Fifth postulate is also independent from the second-order formalization of geometry, while CH is decided in ZFC_2.

Without anticipating the discussion on the similarities and differences between Euclid's Fifth postulate and CH, we have to acknowledge the weakness of using ZFC_2 in an argument in favor of the possibility to decide CH. As a matter of fact, by definition of full second-order logic, every understanding of ZFC_2 calls for a specification of the set-theoretical background. Although the fact that any such specification is enough for deciding CH is a interesting fact, a genuine solution of CH should be univocal, contrary to the possibility of specifying ZFC_2 in different ways.

[20] In Bernays (1967), p. 109.

[21] In this case, the comparison was with Gödel's sentence and the difference was found at the level of decidability with infinitary rules like the ω-rule; see Mostowski (1967).

Hence the possibility to decide CH using second-order logic hardly counts even for an argument in favor of the determinateness of its truth value.

As a consequence, at the London conference in 1965 the problem of deciding the cardinality of the continuum was considered still open, and the need for new axioms was felt by many participants of the discussion. However, different proposals were suggested regarding the methodology to be followed for advancing our knowledge of set theory. Two of them in particular are worth mentioning, as they were presented with clarity and with strong and personal beliefs. On the one hand we can find Kreisel's proposal to use higher-order reasoning as a source of inspiration for first-order axioms.

> Thinking of infinite formulae one is led to find finite formulae (of higher type) which express the meaning of these infinite formulae: this is a principle for getting new axioms, not mentioned by Mostowski. In the present state of affairs this principle is not very practical because (as far as I know) all axioms got in this way can also be got by quite simple reflection principles, but not conversely. In other words, the latter are more fundamental. But the existence of this principle shows clearly that taking second-order notions seriously has at least some consequences for the discovery of new axioms.[22]

We will not discuss the problem of the relationships between second-order logic and set theory, on whose topic there is a wide literature,[23] and the connected subject of the relevance of reflection principles for the justifications of new axioms of set theory[24]—mainly large cardinals axioms. On the other hand we would like to focus our attention on another proposal, by Mostowski, that to our knowledge has not been inspected in detail.

> Models constructed by Gödel and Cohen are important not only for the purely formal reasons that they enable us to obtain independence proofs, but also because they show us various possibilities which are open to us when we want to make more precise the intuitions underlying the notion of a set. Owing to Gödel's work we have a perfectly clear intuition of a set which is predicatively defined by means of a transfinite predicative process. No such clear interpretation has as yet emerged from Cohen's models because we possess as yet no intuition of generic sets; we only understand the relative notion of a set which is generic with respect to a given model. Probably we shall have in the future essentially different intuitive notions of sets just as we have different notions of space, and will base our discussions of sets on axioms which correspond to the kind of sets which we want to study. Although nothing certain can be predicted, we presume that there will be a common part of these various axiomatic systems, and that axioms belonging to this common part will describe the most primitive parts of set-theory which are needed in the expositions of mathematical theories perhaps including the category theory.[25]

We do not consider, as Kreisel suggested, Mostowski's position as a form of skepticism, pointing at the vagueness of the notion of set (opposed to the firm belief that set theory has a clear intended interpretation exemplified by the model(s) constructed

[22]In the discussion of Mostowski (1967), p. 102 of Lakatos (1967).

[23]See Väänänen (2001) and Shapiro (1991) for a sample of this subject.

[24]See Koellner (2009) and Tait (1998) for a sample of this subject.

[25]In Mostowski (1967), p. 94.

by Zermelo for its second-order axiomatization).[26] On the contrary, we believe that it is worth pursuing Mostowski's suggestion: to perform a deeper analysis of the notion of genericity.

It is not clear what an "intuition of generic sets" should be, but we would like to follow the insight according to which a general axiomatic study of the notion of genericity could be useful for a better understanding of the notion of arbitrary set that seems to have a fundamental role in the development of set theory—at least from an historical point of view, as argued in Ferreirós (1999, 2011).

More concretely, at the end of the previous section we argued in favor of the compatibility between the realist component of the method of forcing and the contemporary notion of set shaped in terms of an iterative conception, and we argued that at the root of the latter we find the notion of arbitrary set. What we propose here is a further step that consists in identifying a substantial theoretical link between the notion of arbitrary set and that of generic set. Consequently, we believe that a philosophical analysis of the concept of genericity is not only able to shed light on the realist component of set theory, but it may also make clearer and more formally treatable the notion of arbitrary set.

A good start in this direction consists in trying to specify Bernays' notion of existence of an object "independently of, and prior, to, the construction". It may be reasonably argued that our means of construction are exemplified by first-order logical tools together with the axioms of ZFC—assuming thus a foundational role for set theory. However, when we restrict this notion of construction to a countable transitive model of set theory M, and we consider an infinite set $X \in M$, we have that the collection of all subset of X definable in M clearly does not exhaust $\mathcal{P}(X)$; the power-set of X. Hence, assuming an actual existence of V, we may think of $\mathcal{P}(X)$ as the union of $Def_M(X)$, the collection of $Y \subseteq X$ definable in M, and $Arb_M(X)$, the collection of all other elements of $\mathcal{P}(X)$ that may be thought of as arbitrary when we restrict our means of definitions to M.

However, among the elements of $Arb_M(X)$ there are sets that may be obtained by forcing: subsets of X that belong to generic extensions of M given by filters of posets that belong to M. Starting from these considerations there are many questions we can ask about the relationship between arbitrary sets and generic sets. For example, we could inquire about the possibility of the existence of subsets of X that are purely arbitrary with respect to a given M; that is, whether there exists a $Y \subseteq X$ and an M to which X belongs such that $Y \notin M[G]$ for every M-generic filter G of a poset belonging to M.[27]

---

[26] It is interesting to notice that the possibility of having "essentially different intuitive notions of sets just as we have different notions of space" can be seen as central idea of a multiverse position that we will discuss: that of Hamkins. As we will see in discussing his position, Hamkins moreover argues that the detailed knowledge of the different models of set theory (and consequently of the underlying notions of set) is enough for satisfying our drive for knowledge with respect to the solution of set-theoretical problems as CH.

[27] An example of such a set is $0^\sharp$ that cannot be obtained by forcing. However, the same fact that we may define it makes it non-arbitrary. This simple observation shows that the matter is quite intricate.

We are not interested here in discussing the above question; the above discussion was only meant to illustrate the sense in which we believe that a study of the notion of genericity may help in understanding the notion of arbitrary set.

Before proposing a possible way to carry on Mostowski's proposal—that is to get a more global understanding of the notion of genericity—it is better to given a brief picture of how the study of the possible bifurcation or multifurcation of set theory helped in developing a foundational debate on the nature of the so called multiverse and on the possibility to use formal tools to study such a structure in order to improve our knowledge of the intended interpretation of the axioms of ZFC. In the next section we then present and briefly discuss the different positions of this debate.

12.4 Set Theoretical Multiverses

During the last twenty years the notion of multiverse has been examined in the context of a foundational debate on set theory. Although there is neither a well defined terminology, nor it is clear which are the conceptual boundaries of the notion of multiverse, we will review and discuss the conceptions of Friedman (Friedman and Arrigoni 2013; Friedman et al. 2015; Friedman and Ternullo; Hamkins 2012; Hamkins and Gitman 2010; Väänänen 2001; Woodin 2001, 2011).

For reasons of space, and because some of them are not philosophically structured enough, we choose not to analyze other multiverse conceptions: the one of Shelah (2003), which we may call ZFC-centrism, and mixes a form of naturalism together with a mild formalism; of Magidor (2012), for which some set theories are more equal than others; of Weston (1974, 1976), that applies to a second-order perspective a philosophical interpretation of the independence results obtained by set forcing; and of Steel (Feferman et al. 2000), who intends to analyze the multiverse in order to pursue Gödel's program.

Generally speaking, a multiverse is a collection of structures that express different interpretations of the axioms of set theory and that is closed under specific model-theoretic constructions. Although the method of forcing is an essential ingredient in the construction of a multiverse, the different conceptions vary with respect to which models and methods are allowed. Before entering the discussion of the different notions, it is useful to outline the main terms of distinction and the foundational issues clarified by the different conceptions.

First of all, as we hinted before, which models and tools are admitted is a major element of distinction. Secondly, it is important to understand the aim of the study of the collection of all structures allowed. In particular, to what extent does a given conception aim at studying the multiverse as an independent mathematical structure? And is the outcome of this study to be prescriptive with respect to the axiomatization of set theory? Thirdly, at a more conceptual level, is there a well-defined notion of set that underlies the different models, or is the presence of different models by itself an unavoidable germ of conceptual relativism? Fourth, what is the effect of a multiverse conception on set theoretical truth? And finally, in connection with more ontological issues, does the multiverse conception have any influence on the foundational role of set theory?

We will present the different positions ordering them with respect to an informal notion of quantity relative to the models and tools allowed in the definition of the multiverse, characterizing them by the relevant aspects outlined in the above paragraph.

12.4.1 Platonism

We start by describing Hamkins' view. His multiverse view is very liberal. We may see it as an instance of the widest possible conception, where all set theoretic models and methods are allowed in structuring the mutliverse.

> The background idea of the multiverse, of course, is that there should be a large collection of universes, each a model of (some kind of) set theory. There seems to be no reason to restrict inclusion only to ZFC models, as we can include models of weaker theories ZF, ZF^{-}, KP and so on, perhaps even down to second-order number theory, as this is set-theoretic in a sense. At the same time, there is no reason to consider all universes in the multiverse equally, and we may simply be more interested in the parts of the multiverse consisting of universes satisfying very strong theories, such as ZFC plus large cardinals. The point is that there is little need to draw sharp boundaries as to what counts as a set-theoretic universe, and we may easily regard some universes as more set-theoretic than others.[28]

In this view any set theoretic flavored model should find its place in the multiverse; along with models that, relative to another model, are ill-founded. Hamkins (2012), proposes a strong relativism for what concerns set theoretical notions, married with a strong form of platonism at the meta-theoretical level. Hamkins' notion of existence assimilates his view to the so called full-blooded platonism advocated by Balaguer (1995), for which any possible set theoretic universe of sets has an independent and autonomous existence in the realm of mathematics.

Without entering a detailed discussion of the philosophical strengths and weaknesses of Hamkins' position, we content ourself in pointing out that it is not clear in his view what makes a member of the multiverse a set theoretical structure. The easiest answer would be practice—in accordance with a general naturalism implicit in his ideas—but the vagueness implicit in referring to practice is hardly compatible with a strong platonic philosophy. Moreover, although in Hamkins (2012) (p. 420) we read "[o]n the multiverse view, set theory remains a foundation for the classical mathematical enterprise", it is not clear how to reconcile the different notions of set that are instantiated by the different members of the multiverse, with an ontological foundations for mathematics in set theoretical terms. In other words, how is it possible to compare the different mathematical objects—for example the field of real numbers—living in the different structures? Hamkins is confident in this possibility

[28] Hamkins (2012), p. 439.

since he maintains: "we expect to find all our familiar mathematical objects, such as the integer ring, the real field and our favorite topological spaces, inside any one of the universes of the multiverse".[29] However, a few pages later, Hamkins, criticizing Martin's argument in favor of the categoricity of set theory,[30] points to the difficulty of confronting different concepts of set, acknowledging that " [t]he multiversist objects to Martin's presumption that we are able to compare the two set concepts in a coherent way. Which set concept are we using when undertaking the comparison?".[31] Indeed, we consider the above question problematic also for a view, like Hamkins', that makes the concept of set relative to each model of set theory, but that, nevertheless, assumes a foundational role for set theory. The problem being that if we consider the different models as different foundations of mathematics we lose the possibility to compare them from a conceptual perspective unless we postulate the existence of a meta-theoretical notion of set more primitive than any of the different notions we find in each universe. In other words, if we accept that set theory has a double status of mathematical and foundational theory we should accept that it may able to give a foundation for the mathematics developed in the theory of sets, which includes also the study of the models of ZFC.

Even though it is not clear which is the background theory of sets for this multiverse conception, Hamkins manages to give a formal study of it. Indeed—and this is the major success of his point of view—the extreme freedom of the multiverse is mirrored at a formal level, where many different ways to study this structure are proposed by Hamkins and his collaborators. First of all it is possible to formulate axioms that describe this kind of multiverse—for which a model has been found by Hamkins and Gitman (2010) using the twofold character of a model of ZFC: a model for an external point of view, while a universe for an internal one. Moreover, the application of modal logic to this structure has produced, in Hamkins and Löwe (2008), an interesting analysis of the notion of forcability, similar to the one that Solovay did for provability in Solovay (1976). Finally, the possibility to define, with parameters, the ground model in his generic extension has given rise, in the context of a multiverse structure, of what has been called set theoretic geology[32]: a downward analysis of a multiverse structure.

Even though this multiverse conception has proven itself to be a very fruitful point of view for a formal analysis of the collection of all set theoretical structures, none of these results have been used as arguments in favor of an extension of the ZFC axioms. As a matter of fact, the notion of set theoretical truth underlying this conception is a deflationary notion, and is best exemplified by the next quotation.

> On the multiverse view, consequently, the continuum hypothesis is a settled question; it is incorrect to describe the CH as an open problem. The answer to CH consists of the expansive, detailed knowledge set theorists have gained about the extent to which it holds and fails in

[29] Hamkins (2012), p. 420.

[30] See Martin (2001).

[31] Hamkins (2012), p. 439.

[32] See Reitz (2007), Fuchs et al. (2015) and Hamkins et al. (2008) on this topic.

> the multiverse, about how to achieve it or its negation in combination with other diverse set-theoretic properties.[33]

As a consequence, the notion of set theoretical truth is here assimilated to that of knowledge, in the sense that a precise description of the behavior of CH in the different *known* models of set theory is enough for answering the question about the truth or falsehood of CH.

This aspect of Hamkins' multiverse conception reveals a tension between non-absolute notions at the epistemological level (e.g. what counts as a set theoretical model, or as a solution to the continuum problem) and a strong platonic ontology, where all set theoretical structures have an independent existence and the common mathematical objects are precisely identified across the models.

12.4.2 Conceptualism

The second view we present is the one that, assuming a sufficiently clear picture of the concept of set, considers a multiverse as a collection of set theoretical structures whose study is able to produce new first-order principles intended to formalize aspects of this concept. Of this kind is Friedman's multiverse notion and the related Hyperuniverse Program, as outlined in Friedman and Arrigoni (2013). In Friedman et al. (2015) it is clearly expressed what a hyperuniverse is.

> Our multiverse conception can be described as a multi-level process, which starts with the collection of all transitive countable models of ZFC and ends with the newly identified set-theoretic axioms. We define the collection of all transitive countable models of ZFC as the hyperuniverse. The hyperuniverse occupies the top level in the process.[34]

The mathematical structure given by the hyperuniverse is clearly intended to be analyzed with formal tools and the results thus obtained are meant to be prescriptive for what concerns the choice of new axioms that can augment the ZFC system. Together with this conception of a multiverse comes an informal procedure for the selection of new set theoretical principles that recalls Kreisel's idea of a dialectic relationship between higher-order principles and first-order ones.[35] The starting point for this selection begins with philosophical ideas.

> We have tried to make it clear that our main interest lies in describing philosophical principles which might be turned into higher-order mathematical criteria leading to the selection of universes wherein new set-theoretic axioms, formulated as first-order statements, hold.[36]

Even though the procedure that leads to new axioms is well exemplified in Friedman and Ternullo, it remains open to discussion which are the theoretical

[33] Hamkins (2012), pp. 431-432.

[34] Friedman et al. (2015), p. 14.

[35] See the discussion of Mostowski (1967), pp. 97–103 in Lakatos (1967), for Kreisel's view.

[36] Friedman et al. (2015), p. 19.

reasons for accepting the higher-order principles. This point is crucial, since the proponents of the Hyperuniverse Program intend to give intrinsic reasons for accepting the new first-order axioms produced by the above method. As a consequence, the notion of set underlying Arrigoni-Friedman's multiverse, although very far from both a platonistic position and the belief in the possibility to characterize the concept of set by means of second-order logic, *à la* Kreisler, has, at least at a conceptual level, a stable character that makes it close to a conceptual realism regarding the notion of set.[37]

Quite in tension with this realist component,[38] there is an attention to mathematical practice: "[t]he Hyperuniverse Program may be understood as an attempt to arrive at new *de jure* set-theoretic truths by starting from a picture of the multiverse that faithfully summarizes the full plethora of results obtainable in contemporary set theory",[39] to which the referential set theoretical notion of truth is assimilated.

> ...in formulating the Hyperuniverse Program the expression "true in V" is not used to reflect an ontological state of affairs concerning the universe of all sets as a reality to which existence can be ascribed independently of set-theoretic practice. Instead "true in V" is meant as a *façon de parler* that only conveys information about set-theorists' epistemic attitudes, as a description of the status that certain statements have or are expected to have in set-theorists' eyes.[40]

Instead, the notion of truth proposed in Friedman et al. (2015) is a multilevel one. There is a conceptual level where we find philosophical ideas connected to the concept of set. This is the level of conceptual truth. Then the Hyperuniverse Program intends to specify these ideas with higher-order criteria able to justify first-order principles that act as discriminants for the selection of models of set theory. This is the level of truth of the axioms, whose justification—intrinsic in the intentions of the proponents of the Hyperuniverse Program—depends on the truth of the principles originated by the conceptual level. As a consequence, the truth of the first-order level does not coincide anymore with the idea of true in every model of set theory, but only with being true in the structures that faithfully mirror the higher-order principles that spring from *the* concept of set.

A counterintuitive consequence of this position consists in allowing the possibility of having first-order principles that are logically incompatible, but at the same time both intrinsically justified. However, it is acknowledged by its proponents that the outcomes of the Hyperuniverse Program may influence the lower notion of set theoretical truth. In their words: "However, we also emphasize that the choice of mathematical criteria to select universes is not fixed once and for all and, hence,

[37] Bernays, in describing the weakest form of realism he acknowledges that, in this view, the domain of sets consists of "an ideal projection of a domain of though" (Bernays 1983, p. 261).

[38] Notice that one of the philosophical principles considered in the context of the hyperuniverse program is maximality; also advocated by Gödel.

[39] Friedman and Arrigoni (2013), p. 85.

[40] Friedman and Arrigoni (2013), p.80.

that the statements which we declare as new axioms may vary. This fact leads to a dynamic view of set-theoretic truth".[41]

Nonetheless we have to admit that the above methodology suffers from the theoretical difficulty of shifting the problem of justification of new axioms from the mathematical level to the conceptual one. In other terms, it is not clear how it is possible to justify the philosophical ideas (and the higher-order principles) in terms of the concept of set itself. As a matter of fact, we think that we can find here a tension—similar to the one we described in the case of Hamkins' position—between a static conceptual level and a dynamic one that pertains to mathematical practice; a tension that, in this case, may negatively effect the intrinsic character of the justification sought.

12.4.3 Semantic Realism

The next multiverse position that we would like to present is Woodin's, which perfectly exemplifies a form of realism with respect to truth values of set-theoretical sentences. In other words, sentences like CH, although formally undecidable, are considered to have well determined truth values. Moreover, according to this view, the task of set theorists should consist in discovering the correct solution of every mathematical problem. Another proponent of this form of realism, together with a stronger form of platonism, is Gödel; at least in the late part of his reflections. From this perspective, the study of the multiverse is intended to isolate, among the collection of all allowed structures, the right model (or universe) of set theory, where all set-theoretic problems can find the definitive and right solution. In Woodin's proposal, what needs to be studied is what he calls the *generic multiverse*.

> Let the multiverse (of sets) refer to the collection of possible universes of sets. The truths of set theory according to the multiverse conception of truth are the sentences that hold in each universe of the multiverse. Cohen's method of forcing, which is the fundamental technique for constructing nontrivial extensions of a given (countable) model of ZFC, suggests a natural candidate for a multiverse; the generic multiverse is generated from each universe of the collection by closing under generic extensions (enlargements) and under generic refinements (inner models of a universe of which the given universe is a generic extension).[42]

The study of the generic multiverse is intended to neutralize the effect of forcing, in the search for principles that govern the notion of truth across generic extensions. This study is pursued at the logical level with the definition of a new logic called the Ω-logic. More precisely, Woodin has defined a notion of logical consequence across the generic multiverse in the following way.[43]

---

[41] Friedman et al. (2015), p. 13.

[42] Woodin (2011), p. 102.

[43] We follow here the presentation of Koellner and Woodin (2009).

Definition 12.4.1 (Koellner and Woodin 2009) Suppose there is a proper class of strongly inaccessible cardinals. Suppose T is a theory and φ is a sentence, both in the language of set theory. Let us write

$$T \models_{\Omega} \varphi$$

if whenever $\mathbb{P}$ is a partial order, α is an ordinal, and $G \subseteq \mathbb{P}$ is V-generic, then

$$\text{if } V[G]_{\alpha} \models_{\Omega} T \text{ then } V[G]_{\alpha} \models_{\Omega} \varphi.$$

Moreover, an important result in connection with the above definition consists in showing that the notion of Ω-consequence cannot be changed by forcing; i.e. it does not depends on the model we choose to be the ground model of the generic multiverse.

Theorem 12.4.2 (Koellner and Woodin 2009) *Assume there is a proper class of Woodin cardinals. Suppose T is a theory, φ is a sentence, $\mathbb{P}$ is a partial order and $G \subseteq \mathbb{P}$ is V-generic. Then*

$$V \models \text{“}T \models_{\Omega} \varphi\text{” } \textit{if and only if } V[G] \models \text{“}T \models_{\Omega} \varphi\text{”}.$$

Woodin managed also to define a syntactic counterpart of the Ω-consequence relation, denoted by $\vdash_{\Omega}$.[44] Thanks to these notions, Woodin arrived at formulating a conjecture on the completeness of the Ω-logic, which, if proved correct, would have deep consequences on the notion of truth in set theory (see Larson 2004 and Woodin 1999a). Among other things, it would decide, negatively, the question of the truth of CH. The mathematical results that this multiverse position offers are linked to the so-called Woodin's program.

> One attempts to understand in turn the structures $H(\omega)$, $H(\omega_1)$ and then $H(\omega_2)$. A little more precisely, one seeks to find the relevant axioms for these structures. Since the Continuum Hypothesis concerns the structure of $H(\omega_2)$, any reasonably complete collection of axioms for $H(\omega_2)$ will resolve the Continuum Hypothesis.[45]

Woodin's perspective is often presented in opposition to a multiverse position that accepts the proliferation of the different models of ZFC, and it is normally called the *universe view*. As a matter of fact, this study of the multiverse is intended to find the relevant principles able to describe the true intended model of ZFC. From this perspective, the notion of truth in set theory is considered as fixed. Moreover, the description of the universe V of all sets is a limit towards which the current research in set theory tends. As a consequence, the notion of set that is partially described by the ZFC axiom is not yet completed, but (if ever) it will be so at the end of this process of discovery of set theoretical truths.

---

[44] We will not report here the precise definition of $\vdash_{\Omega}$ due to the non technical character of this work.

[45] Woodin (2001), p. 569.

From this perspective the foundational role of set theory is intended in a very classical—post 1963—way: ZFC partially describes a structure, that our future work will hopefully able to describe in more details, and which will be the context where every mathematical problem will find its definitive solution.

12.4.4 Second-Order Pluralism

We finally present the more abstract multiverse position, due to Väänänen. In line with his higher-order point of view, a multiverse is a collection of universes that needs to be studied from an internal point of view, without resorting to the meta-theoretical study of their connections. Thus the major differences from the previous conceptions consist in considering the members of the multiverse not as models but as *independent* universes and in the proposal of changing the underlying logic to study the multiverse.

> The above thought experiment [used to argue for the presence of sentences that are neither true nor false] shows that allowing a divided reality may call for a re-evaluation of the basic logical operations and laws of logic.[46]

Moreover for Väänänen, as in the case of Hamkins, the study of the multiverse is intended to account for a notion of *absolute undecidability* and not in order to solve the problem of independence.

> Our problem is now obvious: we want two universes in order to account for absolute undecidability and at the same time we want to say that both universes are "everything". We solve this problem by thinking of the domain of set theory as a multiverse of parallel universes, and letting variables of set theory range - intuitively - over each parallel universe simultaneously, as if the multiverse consisted of a Cartesian product of all of its parallel universes. [In note: But the Cartesian product is just a mental image. We cannot form the Cartesian product because we cannot even isolate the universes from each other.] The axioms of the multiverse are just the usual ZFC axioms and everything that we can say about the multiverse is in harmony with the possibility that there is just one universe. [...] The intuition that this paper is trying to follow is that the parallel universes are more or less close to each other and differ only "at the edges". Our multiverse consists of a multitude of universes.[47]

For what concerns the form of realism connected with this multiverse position, we find a notion of existence assimilated to a kind of semantic possibility that transcends a syntactic characterization.

> The idea is not that every model that the axioms of set theory admit is a universe in the multiverse; that would mean that we could dispense with the multiverse entirely and only talk about the axioms.[48]

---

[46] Väänänen (2014), p. 181.
[47] Väänänen (2014), p. 182.
[48] Väänänen (2014), p. 182.

The different universes are for Väänänen developed in "parallel" and any possibility of individuation is banned: "We cannot name individual universes by any means. If we could, we would be able to say what is in that universe and what is not, while at the same time the universes should be "everything"" (Väänänen 2014, p. 190).

As in the discussion of Hamkins's position, the existence of different universes has the effect of relativizing the notion of truth. This relativization has the effect of calling for a different logic in the study of the multiverse. This is indeed the *pars construens* of Väänänen's proposal: the use of a *team semantics*.[49] In some sense this proposal subsumes all the others, because it argues for a general framework where to perform a logical study of possibly different multiverse conceptions.

For what concerns the notion of set, this proposal does not commit itself to any particular one, since it tries to express a general point of view about multiverses. However, the fact that there may be absolutely undecidable propositions makes clear that there are elements of vagueness that may not be eliminated by a subsequent formal investigation. These elements are found in connection with an iterative conception of sets, mainly regarding the notion of the power set operation.

> Problems related to the power-set operation, and more generally to the "arbitrary" sets, is where the multiverse idea emerges. It may just be the nature of the power-set operation that it eludes uniqueness. In anticipation of this we leave the uniqueness untouched and allow different cumulative hierarchies to emerge, in "parallel". Note that this does not mean that we abandon the Power-Set Axiom, which only says that whatever subsets of a given set we happen to have, they can be collected together. [...] What is the use of multiverse set theory if we cannot say anything about the individual universes? The point of the multiverse is that it makes it possible to think of the set theoretic reality as a definite well-defined structure and still doubt the uniqueness of the power-set operation, and keep open the possibility of absolute undecidability[50]

In this sense, although we accept the idea of an intended interpretation, given by a second-order quasi-categoricity result, we still allow that the vagueness of the notion of "arbitrary" set may be responsible for different intended interpretations. However, it remains unspecified in which ways we may specify the notion of arbitrary set, and consequently which are the universes that may count as a legitimate members of the multiverse.

---

[49]See Hodges and Väänänen (2010) and Väänänen (2007) for a presentation of this logic.

[50]Väänänen (2014), p. 191. It is interesting to notice that a similar idea (i.e. the deep link between different specifications of the notion of set and a multiverse perspective) is expressed also by Hamkins. For example in Hamkins (2014) he writes: "Ultimately, the multiverse vision entails an upwardly extensible concept of set, where any current set-theoretic universe may be extended to a much larger, taller universe". But while in the case of Hamkins what is sought is a meta-theoretic perspective that allows him to argue, in Hamkins (2014), that the reason to refute $V = L$ is not its restrictive character (because every countable transitive model can be extended to a model where $V = L$ holds); on the contrary for Väänänen what matters is a internal theoretical perspective that makes him proposing that in order to analyze the multiverse we do not need study meta-theoretical properties of its elements, but instead we just need to change logic.

12.5 Local and Global Perspectives

Although all multiverse positions presented before deal directly—and, in the case of Woodin's, entirely—with the notion of genericity, none of them attempts a direct analysis of this notion. Moreover, even if the study of the multiverse intends to account for a notion of truth that includes also aspects of genericity, this concept is treated locally with respect to a given countable transitive model M. This tension between a local and a global point of view on the matter of independence is perfectly exemplified in the discussion about the independence phenomenon between set theory and geometry.

12.5.1 Set Theory and Geometry

The invention of forcing fostered a comparison between set theory and geometry centered on the discussion of the degrees of independence of CH and of the Parallel Axiom. We would like here to recall the main lines of this debate with the intention of outlining an important aspect that we believe to be instructive also for our analysis of the notion of genericity. We do not pretend to be exhaustive, but only to roughly outline the terms of this discussion.

Going back to the 1965 International Colloquium in the Philosophy of Science, some of the participants proposed a parallel between the independence of CH and the Parallel Axiom. In order to counter an argument in favor of the vagueness of the notion of set and of the same degree of correctness for CH and its negation, Kreisel raised an important objection. It consists in arguing that the comparison between CH and the Parallel Axiom is misleading, because the Parallel axiom has, in his view, a higher degree of independence from the axioms of geometry—being independent from the second-order axiomatization of geometry—than the one of CH with respect to set theory. Indeed, by Zermelo's theorem and the quasi-categoricity of ZFC_2, all second-order models of set theory need to agree on the initial segment of the cumulative hierarchy where the power-set of the natural numbers lies.

Without entering the topic of the possible answers to Kreisel's objection, we believe that the idea expressed by Kreisel consists in proposing a global point of view in set theory in order to partially contain the issues of independence. As a matter of fact, as Weston maintains in Weston (1974)—before criticizing second-order logic as too set theoretical—Kreisel's position presupposes a clear global interpretation of the notion of iterative conception of set.

More recently, this discussion has been resumed by Hamkins (2012) and Koellner (2013) in the attempt to motivate their different foundational perspectives: a liberal multiverse position, proposed by Hamkins, against the universe view, advocated by Woodin and defended in that discussion by Koellner. Hamkins maintains a substantial similarity between the independence of CH and that of the Parallel Axiom, with the intention of depriving CH of a well-determined and unique truth value. On the other hand Koellner—as Weston before him[51]—objects that the phenomenon of independence does not prevent us from asking if the Parallel Axiom is true for physical reality; hence the same question should be asked for CH, with respect to set theoretical reality.

> It makes no more sense to ask whether the parallel pos[t]ulate holds of formal geometry (say, as characterized, by Minimal Geometry) than it does to as whether the axiom of commutativity holds of "group". But it does make sense to ask whether it holds of physical geometry, once we have fixed the bridge equations; this is just the question of whether, given the bridge equations, the geometry selected from the class of formal geometries, satisfies the parallel postulate. The answer for our physical universe (given the bridge equations used in general relativity - sending geodesic to (idealized and uninterrupted) beam of light, etc. -[)] is 'no'.[52]

We agree with Koellner that the independence results are not as conclusive as they may seem at first sight. In the words of Potter: "the independence of the parallel postulate does not in itself show that non-Euclidean geometry describes a way that *space* could be; nor does Cohen's result show on its own that there are two competing theories of *sets*".[53] However, we believe that in order to counter the argument that sees in the independence of CH a signs of the vagueness of the notion of set, there is a better answer than just arguing that a formal proof of the independence of CH (from ZFC) does not satisfy our quest for truth.

The argument we want to propose rests on the recognition of a substantial difference of scope between CH and the Parallel Axiom. We believe that in the discussion on the philosophical meaning of the independence of CH there are two levels that conflate: the local character of the continuum problem and the possibility of having a global notion of set able to characterize the intended interpretation. It is on this particular point where the major difference between the independence of CH and the Parallel Axiom can be found. As a matter of fact, the latter deals with any line and and point, while the former only with a specific and limited kind of objects: sets of natural numbers; or, alternatively, countable ordinals. In other words, the Parallel Axiom is a global principle, while CH is local one.[54]

---

[51] Weston (1974), p. 149.

[52] Koellner (2013), p. 17.

[53] Potter (2004), p. 284.

[54] Of course the local-global character of a sentence depends of the background theory. And indeed also CH may be seen as a global statement in the context of third order arithmetic. However, this fact does not undermine our argument, since a solution to set theoretical problems is often sought in connection with the general notion of set. See for example the following quotation, where Martin and Steel justify PD in terms of large cardinals "Because of their richness and coherence of its consequences, one would like to derive PD itself from more fundamental principles concerning sets in general, principle whose justification is more direct" (Martin and Steel 1989, p. 72). Consequently, if like Gödel we agree that a genuine solution to CH should be in term of a clarification of the general concept of set, then the local-global distinction apply to the discussion on the independence character of CH and the Parallel Axiom. Nonetheless, we personally think that the line of argument that sees the solution of all set-theoretical problem in the clarification of the general concept of set does not do complete justice to the problem of what counts as a genuine solution. But this is a different story about the types of justification of new axioms in set theory.

In our opinion, this distinction of levels answers in a better way the concern about the vagueness of the notion of set raised by the independence of CH because it does not oppose a belief—about a sharp notion of set able to answer the continuum problem—as the quest-for-truth answer does, but it simply removes the general notion of set from the scope of the concern. The reason being that there is a clear difference between the notion of "set belonging to an initial segment of the cumulative hierarchy" and the "iterative conception of set". As a a matter of fact, the former may be characterized as lying in a given set-structure, while the latter cannot. As a consequence, the impossibility to give a decisive answer to CH on the base of the ZFC axioms cannot count as grounds for the belief that "iterative conception of set" is a vague notion, but only that the notion of "subset of the natural number" is not sufficiently specified to give a solution to CH. Of course, the debate on the vagueness or the clarity of the general notion of set may be still open, but what we would like to stress is that the independence of CH may not count as an argument in favor of the vagueness of the more abstract notion of set, but, if ever, only of the more specific one connected with the scope of CH. Quite the contrary, the independence of CH calls for an addition of new principles meant to account for specific instances of the global concept.

A reasonable comparison, instead, should be the one between the Parallel Axiom and the Axiom of Constructibility ($V = L$), as already suggested by Gödel.

> The consistency of the proposition A (that every set is constructible) 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.[55]

For these reasons we believe that the independence of CH does not imply by itself that the underlying notion of set is vague. On the contrary, the fact that the iterative conception is not enough in order to describe the notion of "subset of the natural number" calls for a further analysis of the notion from which the iterative conception originates: that of arbitrary set and, in particular, an analysis of this notion in the context of the interplay between a local and global perspective. A good starting point is exactly the notion of genericity that gave rise to the proliferation of different alternative models of set theory. Moreover, as suggested by Mostowski and by the above discussion about the local and global character of the notions involved, what we propose is to study the multiverse in search for a better global understating of the notion of genericity.

12.5.2 A Generic Forcing

We agree with many participants of the 1965 conference that the moral we can draw from the discovery of the independence of CH is that the ZFC axioms do not entirely capture the notion of set. Consequently, what we need is a better understanding of an

[55] Gödel (1995), p. 155.

idea that is not vague, but whose understanding and formalization are incomplete. Indeed, we do not accept a formalist position like Von Neumann's[56] for which sets are just those mathematical objects whose existence follow form the axioms of ZFC.

In order to study the notion of genericity we embrace a mild form of naturalism: we try to extract philosophical principles from the mathematical use of generic sets, considering mathematical practice as a good source of understanding. In Friedman et al. (2015) a dialectic process is presented between philosophical ideas and mathematical principles. We agree that this dynamic interplay between theoretical ideas and formal principles is very helpful for the progress of mathematics. However we think that even the more philosophically shaped concepts need to be sustained by stable considerations coming from mathematics itself. In a schematic way we can describe our goal as follows: mathematical practice is responsible for the independence of CH and the origin of the notion of genericity; then a formal analysis of this notion should be performed in order to understand the philosophical principles that, sustained by mathematical practice, shape the contemporary notion of set. As a consequence our question is the following.

Question: is the study of the multiverse useful in understanding the notion of generic set and, consequently, of arbitrary set?

A good starting point is the analysis of the notion of genericity not only with respect to sets that are generic for a model, but also with respect to sets that are generic for a multiverse structure.

For these reasons we propose to consider a multiverse as a partial order. Moreover, in order to simplify our study we restrict our analysis to the case of a generic multiverse, as in the case of Woodin. In particular, making this notion more precise, a multiverse is a collection of models that are linked by the relation of generic extension.

Thus we can give the following definition.[57]

Definition 12.5.1 Given a c.t.m. M and a class of forcings Γ, we define the set $\mathbb{M}_\Gamma(M)$ consisting of all generic extensions of M given by elements of Γ. Moreover for $N, N' \in \mathbb{M}_\Gamma(M)$ we set $N' \leqslant_{\mathbb{M}_\Gamma(M)} N$, whenever N' is a generic extension of N by means of a notion of forcing $\mathbb{P} \in \Gamma \cap N$.

In order to show that the above relation is indeed a partial order we need to impose that the trivial forcing,[58] which we may call $\mathbb{T}$, is included in Γ (for reflexivity) and that Γ is closed under products (for transitivity). Notice that, by the definition of generic extension, we also get symmetry without imposing any further condition.

[56]"Here (in the spirit of the axiomatic method) one understands by 'set' nothing but an object of which one knows no more and wants to know no more than what follows about it from the postulates", in Heijenoort (1967), p. 232.

[57]This definition is different both from Viale's category forcing (see Viale 2016) where notions of forcing are considered together with their complete embeddings, instead of models and the relation of generic extension, and from Woodin's $\mathbb{P}_{max}$ (see Woodin 1999b) where generic ultrapowers are considered together with the corresponding elementary embeddings.

[58]By this we mean a forcing that does not produce a proper (generic) extension.

In order to make the above definition useful we need to show that at least for some classes Γ, the forcing of Definition 12.5.1 is non trivial; i.e. forcing with $\mathbb{M}_\Gamma(M)$ over a model N results in a proper extension of N.

Hopefully this is the case, at least for the class $\Gamma_\mathbb{C} = \{\mathbb{C}, \mathbb{T}\}$ containing the Cohen forcing (and the trivial one); i.e. the forcing that allows the addition of a Cohen real. The following observation can be found in Fuchs et al. (2015) and is attributed to Woodin.

Fact 12.5.2 (Woodin) *If W is a countable model of ZFC set theory, then there are forcing extension $W[c]$ and $W[d]$, both obtained by adding a Cohen real, which are non-amalgamable in the sense that there can be no model of ZFC with the same ordinals as W containing both $W[c]$ and $W[d]$.*

As a direct corollary, letting M be a transitive countable model of ZFC and letting $\mathbb{M}_C(M)$ be the notion of forcing consisting of all possible Cohen extensions of M ordered as in Definition 12.5.1, we have the following fact.

Fact 12.5.3 *Forcing with $\mathbb{M}_C(M)$ we obtain a proper generic extension.*

Moreover, in a private conversation between the author and Joel David Hamkins that took place in London in 2011 at the summer school *Set Theory and Higher-Order Logic*, the following theorem has been proved.

Theorem 12.5.4 (Hamkins V.) *Let W be a countable transitive model of ZFC and let $\langle c_n : n \in \omega\rangle$ be a tower of (finitely) mutually W-generic Cohen reals*

$$W \subseteq W[c_0] \subseteq W[c_0][c_1] \subseteq \dots$$

Then there is a W-generic Cohen real d such that $c_n \in W[d]$. As a consequence $W[c_0][c_1]\dots[c_n] \subseteq W[d]$. Furthermore, d is $W[c_0, \dots, c_n]$-generic.

We do not report here the proof of Theorem 12.5.4 because of the non-technical character of this paper, and because we reserve to another, more technical, work a detailed study of these types of forcing. We just outline the relevance of Theorem 12.5.4 for the study of $\mathbb{M}_C(M)$.

Corollary 12.5.5 *The forcing $\mathbb{M}_C(M)$ is σ-closed and so it preserves ω_1.*

The above discussion was meant to show how the study of $\mathbb{M}_C(M)$ may be interesting both from a technical point of view and from a theoretical perspective. Since we intend to analyze elsewhere the technical details of such a study we would like to end this paper pointing at some possible development of this work.

First of all it would be interesting to inquire under which conditions these forcings allow to build new generic models, and in which sense they may be considered new. Such models would show properties that may be considered generically true for a multiverse structure.[59] More generally, it would be interesting to study generic sets with a more complex logical structure, like models. Another very interesting aspect that seems to grow from the study of these forcings of models consists in the possibility of finding sets with different degrees of genericity. Indeed, consider $\mathbb{M}_C(M)$ and the fact that we may meaningly force with it in order to obtain a truly generic extension. Then, the new set so obtained is in some sense more generic with respect to M than the ones obtainable by forcing over M. These are the reasons why we believe that a technical study of these forcings may be able to elucidate the notion of genericity and consequently to give insights into the philosophical problems connected with the use of this technique in set theory.

Acknowledgments This research was financially supported by FAPESP grant number 2013/25095-4 and by the BEPE grant number 2014/25342-4. I would like to thank the organizers and the participants of the conference "SoTFoM III and The Hyperuniverse Programme", where parts of the content of this paper were presented, for the useful discussions. Moreover, I would like to thank Sy David Friedman, David Robert Gilbert, Joel David Hamkins, Geoffrey Hellman, Adrian Richard David Mathias, Boban Veličković and Matteo Viale for the instructive discussions and contributions that helped me shaping this paper. Finally I would like to thank two anonymous referees for their useful comments.

References

Balaguer, M. (1995). A platonist epistemology. *Synthese*, *103*, 303–325.

Bell, J. (2005). *Set theory. Boolean valued models and independence proofs*. Oxford: Oxford Science Publications.

Bernays, P. (1967). What do some recent results in set theory suggest? In I. Lakatos (Ed.), *Problems in the philosophy of mathematics* (pp. 109–112). Amsterdam: North-Holland.

Bernays, P. (1983). On mathematical platonism. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics: Selected readings* (pp. 258–271). Cambridge: Cambridge University Press.

Cohen, J. P. (1963). The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *50*(6), 1143–1148.

Feferman, S., Friedman, H., Maddy, P., & Steel, J. (2000). Does mathematics need new axioms? *Bulletin of Symbolic Logic*, *6*(4), 401–446.

Ferreirós, J. (1999). *Labyrinth of thought. A history of set theory and its role in modern mathematics*. Basel: Birckhäuser.

Ferreirós, J. (2011). On arbitrary sets and ZFC. *Bullettin of Symbolic Logic*, *17*(3), 361–393.

Friedman, S. D. (2000). *Fine structure and class forcing*. Berlin: Walter de Gruyter.

Friedman, S. D., Antos, C., Honzik, R., & Ternullo, C. (2015). Multiverse conceptions in set theory. *Synthese*, *192*(8), 2463–2488.

Friedman, S. D., & Arrigoni, T. (2013). The hyperuniverse program. *Bulletin of Symbolic Logic*, *19*(1), 77–96.

Friedman, S. D., & Ternullo, C. (2014). Believing the new axioms. Preprint available at http://www.logic.univie.ac.at/sdf/papers/joint.claudio.pdf

[59]It could be interesting to study these models in comparison to what in Shelah (2003), calls a *typical* model; this notion is also relevant for Friedman's multiverse conception. Another interesting comparison would be with Robinson's application of forcing to model theory; see Robinson (1971).

Fuchs, G., Hamkins, J. D., & Reitz, J. (2015). Set-theoretic geology. *Annals of Pure and Applied Logic*, *166*(4), 464–501.

Gödel, K. (1995). *Collected works. Volume III: Unpublished essays and lectures*. Oxford: Oxford University Press.

Hamkins, J. (2012). The set-theoretic multiverse. *Review of Symbolic Logic*, *5*, 416–449.

Hamkins, J., & Gitman, V. (2010). A natural model of the multiverse axioms. *Notre Dame Journal Formal Logic*, *51*(4), 475–484.

Hamkins, J., & Löwe, B. (2008). The modal logic of forcing. *Transactions of the American Mathematical Society*, *360*, 1739–1817.

Hamkins, J., Reitz, J., & Woodin, H. (2008). The ground axiom is consistent with $V \neq HOD$. *Proceedings of the AMS*, *136*, 2943–2949.

Hamkins, J. D. (2014). A multiverse perspective on the axiom of constructiblity. In T. A. Slaman, C. Chong, Q. Feng, & W. H. Woodin (Eds.), *Infinity and truth* (pp. 25–46). Hackensack: World Scientific.

Hamkins, J. D., & Seabold, E. (2012). Well-founded Boolean ultrapowers as large cardinal embeddings. Preprint available at http://arxiv.org/abs/1206.6075

Hodges, W., & Väänänen, J. (2010). Dependence of variables construed as an atomic formula. *Annals of Pure and Applied Logic*, *161*(6), 817–828.

Jech, T. (2002). *Set theory, the third millennium edition, revised and expanded*. Berlin: Springer.

Koellner, P. (2009). On reflection principles. *Annals of Pure and Applied Logic*, *157*(2–4), 206–219.

Koellner, P. (2013). *Hamkins on the multiverse*. Unpublished notes.

Koellner, P., & Woodin, H. (2009). Incompatible ω-complete theories. *Journal of Symbolic Logic*, *74*(4), 1155–1170.

Kunen, K. (1980). *Set theory. An introduction to independence proofs*. Amsterdam: North-Holland.

Lakatos, I. (1967). *Problems in the philosophy of mathematics*. Amsterdam: North-Holland.

Larson, P. B. (2004). *The stationary tower: Notes on a course by W. Hugh Woodin*. Providence: AMS.

Magidor, M. (2012). Some set theories are more equal then others. Notes from the EFI Project Conference.

Martin, D. (2001). Multiple universes of sets and indeterminate truth values. *Topoi*, *20*(1), 5–16.

Martin, D., & Steel, J. (1989). A proof of projective determinacy. *Journal of the American Mathematical Society*, *2*(1), 71–125.

Mostowski, A. (1967). Recent results in set theory. In I. Lakatos (Ed.), *Problems in the philosophy of mathematics* (pp. 82–95). Amsterdam: North-Holland.

Potter, M. (2004). *Set theory and its philosophy*. Oxford: Oxford University Press.

Reitz, J. (2007). The ground axiom. *Journal of Symbolic Logic*, *72*(4), 1299–1317.

Robinson, A. (1971). Forcing in model theory. *Symposia Mathematica*, *V*, 69–82.

Shapiro, S. (1991). *Foundations without foundationalism. A case for second-order logic*. Oxford: Oxford University Press.

Shelah, S. (2003). Logical dreams. *The Bulletin of the American Mathematical Society*, *40*(2), 203–228.

Solovay, R. (1976). Provability interpretations of modal logic. *Israel Journal of Mathematics*, *25*(3–4), 287–304.

Tait, W. W. (1998). Zermelo's conception of set theory and reflection principles. In M. Schirn (Ed.), *The philosophy of mathematics today* (pp. 469–483). New York: Clarendon Press.

Van Heijenoort, J. (1967). *From Frege to Gödel: A source book in mathematical logic, 1879–1931*. Cambridge: Harvard University Press.

Väänänen, J. (2001). Second order logic and foundations of mathematics. *The Bulletin of Symbolic Logic*, *7*(5), 504–520.

Väänänen, J. (2007). *Dependence logic*. Cambridge: Cambridge University Press.

Väänänen, J. (2014). Multiverse set theory and absolutely undecidable propositions. In J. Kennedy (Ed.), *Handbook of set theory* (pp. 180–208). Cambridge: Cambridge University Press.

Viale, M. (2016). Category forcings, MM^{+++}, and generic absoluteness for the theory of strong forcing axioms. *Journal of the American Mathematical Society, 29*(3), 675–728.

Weston, T. (1974). The continuum hypothesis: Independence and truth-value. PhD thesis.

Weston, T. (1976). Kreisel, the continuum hypothesis and second order set theory. *Journal of Philosophical Logic*, *5*(2), 281–299.

Woodin, H. (1999a). *The axiom of determinacy, forcing axioms, and the nonstationary ideal*. Berlin: Walter de Gruyter.

Woodin, H. (1999b). *The axiom of determinacy, forcing axioms, and the nonstationary ideal*. Berlin: Walter de Gruyter and Co.

Woodin, H. (2001). The continuum hypothesis. Part I. *Notices of the American Mathematical Society*, *48*(6), 567–576.

Woodin, H. (2011). The realm of the infinite. In M. Heller & W. H. Woodin (Eds.), *Infinity. New research frontiers* (pp. 89–118). Cambridge: Cambridge University Press.

Zermelo, E. (1932). *Georg Cantor. Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. New York: Springer.

Zermelo, E. (2010). *Collected works* (Vol. I). New York: Springer.

Part III
The Logic Behind Mathematics: Proof, Truth, and Formal Analysis

Chapter 13
What's so Special About the Gödel Sentence $\mathcal{G}$?

Mario Piazza and Gabriele Pulcini

Abstract The very fact that the Gödel sentence $\mathcal{G}$ is independent of Peano Arithmetic fuels controversy over our access to the *truth* of $\mathcal{G}$. In particular, does the truth of $\mathcal{G}$ (of the form $\forall x \varphi(x)$) precede the truth of its numerical instances $\varphi(0)$, $\varphi(1)$, $\varphi(2)$, ..., as the so-called standard argument induces one to believe? This paper offers a shift in perspective on this old problem. We start by reassessing Michael Dummett's 1963 argument which seems to speak in favour of the priority of the truth of the numerical instances of $\mathcal{G}$ over the truth of $\mathcal{G}$ itself. In opposition to some recent criticisms of Dummett's argument, we argue that the latter is not reducible to the standard one. We then point out its *prototypical* nature in the sense individuated by Jacques Herbrand. This shift in perspective brings us to the claim that the controversy over the priority between $\mathcal{G}$ and its numerical instances endures only because the problem is ultimately ill-posed. An encompassing moral about the epistemological mechanism of prototype proofs is also drawn.

Keywords Incompleteness theorems · Standard and non-standard arguments for the truth of the Gödel sentence · Prototype proofs

13.1 Introduction

Like the Prince of Denmark in *Hamlet*, the Gödel sentence $\mathcal{G}$ is the main character in the story of the incompleteness of Peano Arithmetic (PA henceforth). As the main character of the story, however, $\mathcal{G}$ is decidedly *sui generis*: it is precisely because $\mathcal{G}$ is independent of PA—i.e. $\mathsf{PA} \nvdash \mathcal{G}$ and $\mathsf{PA} \nvdash \neg\mathcal{G}$—that the threads of its truth are

M. Piazza (✉)
Department of Philosophy, University of Chieti-Pescara, Chieti, Italy
e-mail: mpiazza@unich.it

G. Pulcini
Centre for Logic, Epistemology and History of Science,
State University of Campinas, Campinas, Brazil
e-mail: gab.pulcini@cle.unicamp.br

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_13

hard to trace and untangle. Our concern in this paper is the problem of the extent and nature of the reasoning that leads to the recognition of the truth of $\mathcal{G}$.

It is well-known that if PA is a consistent theory, then $\mathcal{G}$ is *true-in-the standard model* $\mathcal{N}$, so that its independence may be diagnosed as "not a radical one" unlike that of the continuum hypothesis (Gaifman 2000, pp. 462–63). But the interest of the truth of $\mathcal{G}$ is largely in the epistemic process of its 'objectivation', which corresponds to the deductive path by which that truth can be ascertained. It is fair to say that an explanation of our access to the truth of $\mathcal{G}$ via this path is a task that any account of the significance of the incompleteness phenomenon must discharge. And of course, this explanation also touches upon the contentious area of the nature of $\mathcal{G}$ itself.

In effect the status of the sentence $\mathcal{G}$ is the object of considerable controversy because it is ambiguous between a *metatheoretical* and an *arithmetical* meaning. Unsurprisingly, such ambiguity is intrinsic to the Gödelian construction of $\mathcal{G}$ as a (provable) fixed point of the predicate $\neg\mathsf{Prov}(y)$, where $\mathsf{Prov}(y)$—the *provability* predicate—is in turn defined as $\exists x \mathsf{Prf}(x, y)$. Thus, on the one hand, $\mathsf{Prf}(x, y)$ is assigned a metatheoretical meaning, insofar as PA proves $\mathsf{Prf}(\overline{n}, \overline{m})$ if and only if n is the Gödelian code of a PA-proof ending with a formula whose code is m. And on the other, $\mathcal{G}$ is provably equivalent to the sentence $\forall x \neg\mathsf{Prf}(x, \overline{\ulcorner \mathcal{G} \urcorner})$, where $\mathsf{Prf}(x, y)$ is a decidable predicate, i.e. there exists a total recursive function $\rho(x, y)$ such that $\rho(n, m) = 1$, if n can be decoded into a PA-proof of the sentence encoded by m, and $\rho(n, m) = 0$, otherwise. Let us call $\mathcal{A}$ and $\mathcal{A}^{-1}$ the encoding and decoding algorithms which govern the procedures of encoding and decoding, respectively. More specifically, $\rho(n, m)$ returns its output by performing the following two steps:

(i) it applies $\mathcal{A}^{-1}$ to n so as to (possibly) get a finite sequence of natural numbers $n_1, \ldots, n_i$[1];
(ii) it confronts n_i with m.

If $n_i = m$, then $\rho(n, m) = 1$; otherwise, $\rho(n, m) = 0$. Now, it is possible to skip any metatheoretical commitment by shifting the focus away from the predicate $\mathsf{Prf}(x, y)$ to the decoding algorithm $\mathcal{A}^{-1}$ and so to the arithmetical function $\rho(x, y)$. Let $\ulcorner \mathcal{G} \urcorner = g$; according to this view, $\mathcal{G}$ is speaking of the function $\rho(x, y)$ by saying that, for all $n \in \mathbb{N}$, $\rho(n, g) = 0$. Although $\mathcal{A}$ describes an effective procedure allowing the numeralwise representability of proofs and formulas, the function $\rho(x, y)$ needs only to apply the reverse algorithm $\mathcal{A}^{-1}$ for decoding a natural number into a sequence of natural numbers and *no metatheoretical feature seems to come into play*.

However, the technical fact that the metatheoretical and the arithmetical readings turn out to be equivalent modulo the numeralwise representability of proofs and formulas matters less than the recognition of their conceptual difference: according to the metatheoretical reading, $\mathcal{G}$ amounts to a self-referential sentence asserting that "$\mathcal{G}$ is not provable within PA". By contrast, the arithmetical reading tells us that $\mathcal{G}$ amounts to a sentence asserting a well-established property about a certain encoding algorithm satisfying certain properties.

---

[1] This is the sequence encoding the PA-proof $\alpha_1, \ldots, \alpha_n$ of α_n where $\ulcorner \alpha_i \urcorner = n_i$.

Twentieth-century philosophy of logic has proposed two tactics for establishing the truth of $\mathcal{G}$, two tactics mirroring the two sides of $\mathcal{G}$'s ambiguity: the 'standard' argument and the 'non-standard' one. More specifically, these arguments correspond to two opposite ways of looking at the relationship between the truth of

$$\mathcal{G}(\leftrightarrow \forall x \neg \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner))$$

and that of its numerical instances:

$$\neg \mathsf{Prf}(0, \ulcorner \mathcal{G} \urcorner), \neg \mathsf{Prf}(1, \ulcorner \mathcal{G} \urcorner), \neg \mathsf{Prf}(2, \ulcorner \mathcal{G} \urcorner), \ldots$$

The *naive* standard argument—epitomized by most textbook demonstrations—somehow manages to secure a metatheoretical niche for the sentence $\mathcal{G}$: in a nutshell, $\mathcal{G}$ says 'I am unprovable in PA', the First Incompleteness Theorem establishes the unprovability of $\mathcal{G}$ within PA, thence it *ipso facto* verifies $\mathcal{G}$. We call this argument 'naive' in order to distinguish it from its formal renditions upon which we shall enlarge *infra*. Once the truth of $\mathcal{G}$ in $\mathcal{N}$ has been reached in this way, the truth of its numerical instances can be nonchalantly derived by universal specification:

$$\frac{\forall x \neg \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner) \text{ is true}}{\neg \mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner) \text{ is true}} \, .$$

The non-standard argument for the truth of $\mathcal{G}$ was inaugurated by Michael Dummett in his 1963 essay *The Philosophical Significance of Gödel's Theorem* (Dummett 1963). On this view, we are invited to follow a deductive path from the truth of the numerical instances of $\mathcal{G}$ to the truth of $\mathcal{G}$. When the truth of each single numerical instance is verified, an application of the Tarskian biconditional for the universal quantification—or, equivalently, a step of the ω-rule (Piazza and Pulcini 2015)—leads us to conclude that $\mathcal{G}$ is true in the standard model $\mathcal{N}$.

Unlike the non-standard argument, the standard one (either naive or formal) makes it tempting to believe that $\mathcal{G}$ is a *special* sentence, inasmuch as the ascertainment of its truth as a universally quantified sentence is not guided by our thinking about the truth of its instances. As a recent defender of this view puts it, $\mathcal{G}$ is "*very special in that it is (equivalent to) a universal sentence which is epistemologically prior to its numerical instances*" (Serény 2011, p. 48). The argument to the effect that $\mathcal{G}$ is not a genuinely arithmetical proposition may be presented as a *modus tollens*: if $\mathcal{G}$ were a genuinely arithmetical sentence, a proof of the truth of $\mathcal{G}$ in $\mathcal{N}$ would be available to us, a proof in which the numerical instances of $\mathcal{G}$ would be prior to $\mathcal{G}$. But this proof is not available, since non-standard arguments are nothing other than standard arguments in disguise (Serény 2011, *ibidem*).

Now, it seems plausible to say that in a *usual* sort of proof of the number-theoretical sentence $\psi \equiv \forall x \varphi(x)$, the recognition of the truth of each one of the instances of ψ is *preliminary* to that of the truth of ψ (Serény 2011, p. 48). Yet, 'usual sort' in 'usual sort of proof' is vague. Certainly, there is a clear sense in which the property ψ in any *inductive* proof of $\forall x \psi(x)$ is shown to be recursively propagated from one natural number to its successor so as to cover the totality of the natural numbers. Namely, we feel that an inductive proof of $\forall x \psi(x)$ has a temporal dimension imposed by its recursive structure: firstly we prove that $\psi(0)$ holds true, then that $\psi(1)$ holds true as well, and so on. The passage to the limit upon such an infinite temporal sequence authorizes us to conclude $\forall x \psi(x)$. Therefore, there can be no question of what the epistemological priority means in this case: the truth of the numerical instances $\psi(0), \psi(1), \ldots$ has algorithmically to precede the achievement of the truth of $\forall x \psi(x)$. Nevertheless, proofs by induction are not the whole of the story about proofs of universal quantified sentences in arithmetic. This consideration, simple as it is, will be fundamental to our discussion.

In this paper, we propose a reassessment of the interplay between the standard and non-standard view of the truth of $\mathcal{G}$, a reassessment which aims at constituting a corrective for some widespread misunderstandings in the literature on the topic, and, in particular, commits us to denying a trio of claims made by Serény which fit orthodoxy: (i) standard strategies represent the only avenue to the truth of $\mathcal{G}$; (ii) non-standard arguments are formal standard arguments in disguise; therefore, (iii) $\mathcal{G}$ is not a genuinely arithmetical proposition, but rather a metatheoretical 'artifact'.

The naive standard view is both venerable and, at least initially, candid, but it embroils us in acute philosophical difficulties. There are indeed two principal reasons why its metatheoretical baggage provides no guidance as to the logical nature of the proofs leading to the truth of $\mathcal{G}$. The first reason is that the assignment of a metatheoretical meaning to $\mathcal{G}$ seems to be compromising: a cloud of circularity hangs over an investigation devoted precisely to evaluating the genuine metatheoretical content of $\mathcal{G}$, while there is nothing intrinsic to $\mathcal{G}$ which demands, or even encourages, a metatheoretical treatment. Moreover, as we have said, the naive standard argument thinks of $\mathcal{G}$ in metatheoretical terms, whereas non-standard arguments regard it as an ordinary arithmetical sentence. But the two lines of reasoning run conceptually in parallel, so that a wrong way to argue is to assimilate one to the other by making one interpretation prevail over the other. We claim that the best chance for a strategy for addressing the problem lies in framing the naive standard and non-standard arguments into a common *formal* setting in which answers about the two kinds of argument can be clearly stated. This is why the non-standard argument will be compared with formal renditions of the naive standard argument. Needless to say, in logic the devil is always in the detail.

The second consideration against the metatheoretical interpretation is that it is philosophically controversial in making great play of the diagonalization lemma which states that $\mathcal{G} \leftrightarrow \neg\mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$ is a theorem of PA. What are we to make of it? This lemma indeed collapses into the claim that $\mathcal{G}$ and $\neg\mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$ are mutually interchangeable with regard to the provability in PA: $\mathsf{PA} \vdash \mathcal{G} \leftrightarrow \neg\mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$. In this way, any intensional meaning attached to the sentence $\mathcal{G}$ which differs from interchangeability stands in need of further support. As Volker Halbach and Albert Visser note:

> The only 'self-reference-like' feature of the Gödel sentence [$\mathcal{G}$] that is used in the usual proofs of Gödel's first incompleteness theorem is the derivability of the equivalence [$\mathcal{G} \leftrightarrow \neg\mathsf{Prov}(\ulcorner\mathcal{G}\urcorner)$]; in other words, the only feature needed is the fact that [$\mathcal{G}$] is a (provable) fixed point of [$\neg\mathsf{Prov}(\ulcorner\mathcal{G}\urcorner)$]. But the fact that a sentence is a fixed point of a certain formula expressing a certain property does by no means guarantee that the sentence ascribes that property to itself […]; and whether [$\mathcal{G}$] is also self-referential, or 'states its own unprovability' in any sense whatsoever is not relevant for proof of the first incompleteness theorem. (Halbach and Visser 2014a, pp. 671–672)

On the other hand, we think that the Dummettian non-standard approach points towards something pivotal to the very incompleteness phenomenon, namely the constraint of consistency. Nevertheless, it fails to disengage the matter of the arithmetical status of $\mathcal{G}$ from the misleading relation of priority between $\mathcal{G}$ and its instance, a relation which it implicitly assumes. According to the account that we shall recommend in this paper, the issue of priority on which the sides in the controversy disagree is ultimately a *Scheinstreit*. Once it is properly acknowledged that non-standard arguments for the truth of $\mathcal{G}$ have features which mark a type of *arithmetical* reasoning, a reasoning constitutively involving *genericity*, it should also become apparent that we cannot hive off $\mathcal{G}$ from its numerical instances. Hence the worry about priority becomes critically misguided. Indirectly, our argumentation is thus meant also to provide negative evidence for Isaacson's conjecture, whereby if we are to give a proof of *any* true PA sentence which is independent of PA, then we will need to appeal to ideas that go beyond those that are required in understanding PA (Isaacson 1987).

Roadmap: in the next section, we briefly recall some basic facts about Δ_0- and Σ_1-completeness that we will use in the sequel. In Sect. 13.3, we give Dummett's argument an unhurried examination by stressing its critical junctures. In Sect. 13.4, we argue—contrary to Serény's contention—that it is the standard demonstration of the truth of $\mathcal{G}$ which proves to be a 'zipped' version of Dummett-inspired arguments. In Sect. 13.5, we widen the appeal of non-standard arguments by pointing out their *prototypical* nature, so that we are not entitled to take $\mathcal{G}$ to be a *sui generis* sentence. In Sect. 13.6, we briefly take stock.

13.2 Preliminaries: Δ_0- and Σ_1-Completeness

Definition 1 (*Logical complexity*)

- A formula α belongs to the set Δ_0 if it is equivalent to a closed formula α' in which all the quantifiers, if any, are bounded.
- A formula α belongs to Σ_1 (resp. Π_1) if it is equivalent to a closed formula $\alpha' \equiv \exists x \beta(x)$ (resp. $\alpha' \equiv \forall x \beta(x)$) such that $\beta[t/x] \in \Delta_0$.
- A formula α belongs to Σ_{n+1} (resp. Π_{n+1}) if it is equivalent to a closed formula $\alpha' \equiv \exists x \beta(x)$ (resp. $\alpha' \equiv \forall x \beta(x)$) such that $\beta[t/x] \in \Pi_n$ (resp. $\beta[t/x] \in \Sigma_n$).

Example 13.2.1 Both the Gödelian propositions $\mathcal{G}$ and $Cons(\mathsf{PA})$ are Π_1-statements.

Remark 1 Whereas $\alpha \in \Sigma_n$ if, and only if, $\neg\alpha \in \Pi_n$, the set of Δ_0 formulas is closed under negation.

Proposition 13.2.1 *Let t, s be two closed arithmetical terms:*

(1) *if* $\mathcal{N} \vDash t = s$, *then* $\mathsf{PA} \vdash t = s$,
(2) *if* $\mathcal{N} \vDash t \neq s$, *then* $\mathsf{PA} \vdash t \neq s$,
(3) $\mathsf{PA} \vdash \overline{n} \leqslant \overline{m} \rightarrow (\overline{n} = \overline{0} \vee \overline{n} = \overline{1} \vee \ldots \vee \overline{n} = \overline{m})$.

Proof The reader can find all the proofs in Rautenberg (2000). □

Theorem 13.2.2 (Δ_0-decidability) *If α is a closed Δ_0-formula, then either* $\mathsf{PA} \vdash \alpha$ *or* $\mathsf{PA} \vdash \neg\alpha$.

Proof Let $\alpha \in \Delta_0$; we proceed by induction on the number of logical connectives occurring in α.

<u>*Base*</u>. If no logical connective occurs in α, then $\alpha \equiv t = s$ with t, s closed terms. It is either $\mathcal{N} \vDash t = s$ or $\mathcal{N} \vDash t \neq s$ and so Proposition 13.2.1 gives us the base case.

<u>*Step*</u>. Proposition 13.2.1 enables us to stress the following conversions

$\exists x \leqslant k\alpha(x) \Leftrightarrow \alpha(0) \vee \alpha(1) \vee \ldots \vee \alpha(k)$
$\forall x \leqslant k\alpha(x) \Leftrightarrow \alpha(0) \wedge \alpha(1) \wedge \ldots \wedge \alpha(k)$,

for turning any quantified Δ_0-formula into an equivalent one without quantifiers. Then it is easy to see that any Boolean composition of decidable propositions is, in turn, decidable. □

Corollary 13.2.3 (Δ_0-completeness) *For any closed $\alpha \in \Delta_0$, if $\mathcal{N} \vDash \alpha$, then* $\mathsf{PA} \vdash \alpha$.

Proof Let $\mathcal{N} \vDash \alpha$, but $\mathsf{PA} \nvdash \alpha$. Since $\alpha \in \Delta_0$, by Theorem 13.2.2, it would be $\mathsf{PA} \vdash \neg\alpha$ against the soundness of PA w.r.t. $\mathcal{N}$. □

Lemma 13.2.4 PA *is Σ_1-complete w.r.t. $\mathcal{N}$ if, and only if, it is Δ_0-decidable.*

Proof ($\Rightarrow$) Let $\mathsf{PA} \nvdash \alpha$, with α closed and in Δ_0. By Σ_1-completeness, we obtain $\mathcal{N} \nvDash \alpha$ and so $\mathcal{N} \vDash \neg\alpha$. Since $\neg\alpha \in \Delta_0$, we perform a further step of Σ_1-completeness so as to obtain $\mathsf{PA} \vdash \alpha$.

($\Leftarrow$) Let $\exists x\alpha(x)$ be a closed Σ_1-formula such that $\mathcal{N} \vDash \exists x\alpha(x)$. By definition, there is an $n \in \mathbb{N}$ for which $\alpha(n)$ is true-in-$\mathcal{N}$. Then, we can apply the Δ_0-completeness to get $\mathsf{PA} \vdash \alpha(\overline{n})$ and so conclude $\mathsf{PA} \vdash \exists x\alpha(x)$. □

Theorem 13.2.5 (Σ_1-completeness) PA *is Σ_1-complete w.r.t. $\mathcal{N}$.*

Proof Straightforwardly by Theorem 13.2.2 and Lemma 13.2.4. □

Corollary 13.2.6 *If $\alpha \in \Pi_1$ is independent of* PA, *then $\mathcal{N} \vDash \alpha$. In particular, we have that $\mathcal{N} \vDash \mathcal{G}$ and $\mathcal{N} \vDash Cons(\mathsf{PA})$.*

Proof By Σ_1-completeness, we obtain $\mathcal{N} \nvDash \neg\alpha$ from $\nvdash_{\mathsf{PA}} \neg\alpha$, and so $\mathcal{N} \vDash \alpha$. Both the Gödelian propositions $\mathcal{G}$ and $Cons(\mathsf{PA})$ instantiate the case just explained so that $\mathcal{N} \vDash \mathcal{G}$ and $\mathcal{N} \vDash Cons(\mathsf{PA})$. □

13.3 Unfolding Dummett's Argument

Dummett's 1963 essay suggests a way of recognizing the truth of $\mathcal{G}$ by means of the truth of its numerical instances. His proposal is contained in a highly compressed passage and, admittedly, it needs refinement in order to make it fully determinate. For the sake of perspicuity, we unfold Dummett's argumentation by breaking it down into five steps. Steps 1 and 2 sketch a non-standard route to the conviction that $\mathcal{G}$ is true, whilst Steps 3, 4 and 5 provide some methodological observations on that route.

> Step 1: *The particular predicate $A(x)$ is such that, if $A(n)$ is false for some numeral n, then we can construct a proof in the system of $\forall x A(x)$.* (Dummett 1963, p. 192)

By 'predicate $A(x)$', Dummett clearly means the formula $\neg\mathsf{Prf}(x, \ulcorner\mathcal{G}\urcorner)$. So, his point can be glossed by saying that both $\neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ and $\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ are Δ_0-formulas and that PA is Δ_0-complete. That is, for any Δ_0-formula α:[2]

$$\mathsf{PA} \vdash \alpha \Leftrightarrow \mathcal{N} \models \alpha.$$

The idea here is that if $\neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ is false-in-$\mathcal{N}$, then $\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ is true-in-$\mathcal{N}$ and so $\mathsf{PA} \vdash \mathsf{Prf}(\overline{n}, \overline{\ulcorner\mathcal{G}\urcorner})$. By the very definition of the Gödelian provability predicate, the natural number n could be recursively decoded into a PA-proof of $\mathcal{G}$, that is the sentence $\forall x \neg\mathsf{Prf}(x, \overline{\ulcorner\mathcal{G}\urcorner})$ would be provable within PA.

> Step 2: *From this it follows — on the hypothesis that the system is consistent — that each $A(0), A(1), A(2), \ldots$ is true.* (Dummett 1963, p. 192)

By Step 1, we know that if $A(n) \equiv \neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ were false, then $\mathcal{G}$ would be provable within PA. However, from the First Incompleteness Theorem, we also know that, if PA is consistent, then $\mathcal{G}$ is independent of PA. So, *under the hypothesis of the consistency of* PA: for all n, $A(n)$ is true.

[2]Dummett's argument is originally developed by carefully avoiding any reference to the notion of (standard) model. Anyhow, we are stressing the model-theoretic machinery in view of a common formal ground for expressing the standard and non-standard proofs and then comparing them. Dummett notoriously attacks the following line of reasoning:

> Under bivalence, one has to choose between $\mathcal{G}$ and $\neg\mathcal{G}$. But $\mathcal{G}$ is verified by the standard model, whereas only non-standard models verify its negation $\neg\mathcal{G}$. Now, inasmuch as $\mathcal{N}$ is the *intended* model of PA—that is, the structure which is supposed to deliver our basic intuitions about naturals—$\mathcal{N}$ also expresses the context in which formal statements are actually verified or falsified. Which is to say, the privileged point of view whereby, so to say, truth is really true and false is really false.

The basic problem with this argument is that, instead of deciding which of the two sentences $\mathcal{G}$ and $\neg\mathcal{G}$ is the true one, one actually chooses between standard and non-standard models. Dummett correctly cautions us not to conflate the *explanandum*, i.e. the structure of naturals represented by $\mathcal{N}$, with the *explanans*, i.e. the formal theory PA which is supposed to provide a formal ground for our pre-formal intuitions about natural numbers. In conclusion—Dummett says—the truth of $\mathcal{G}$ is returned by a process which is conceptually circular.

> Step 3: [...] *the transition from saying that all the statements* $A(0), A(1), A(2), \ldots$ *are true to saying that* $\forall x A(x)$ *is true is trivial. The principle of reasoning not embodied in the system, which we employ in arriving at the truth of* $\forall x A(x)$, *is not this transition, but rather that which leads us to assert that all of the statements* $A(0), A(1), A(2), \ldots$ *are true*. (Dummett 1963, p. 192)

In contrast to what one might think, the critical deductive passage is not that from the numerical instances to the universal quantification, namely the transition which is not formalizable within PA.[3] Rather, Dummett is telling us that it is the phase concerning the verification of the numerical instance that is the critical one, inasmuch as consistency is *needed*:

> Step 4: *Therefore it is in the reasoning which shows that all of* $A(0), A(1), A(2), \ldots$ *are true, and not in the quite evident step from there to the truth of* $\forall x A(x)$, *that we have to appeal to something which cannot be formalized in the system: namely to the argument which is intended to show that the system possesses overall consistency.* (Dummett 1963, p. 192)

What makes the process of verification of all the numerical instances not trivial is precisely the fact that it is understood as constrained by consistency. Indeed, as we have observed with regard to Steps 1 and 2, the verification process is centred on the independence of $\mathcal{G}$ which is established by the First Incompleteness Theorem *under the hypothesis of the consistency of* PA.

Dummett continues:

> Step 5: *Considered as an argument to a hypothetical conclusion — that if the system is consistent, then* $\forall x A(x)$ *is true — this reasoning can of course be formalized in the system.* (Dummett 1963, p. 192)

The observation in Step 4 is reinforced by looking at the Second Incompleteness Theorem which states that

$$\mathsf{PA} \vdash Cons(\mathsf{PA}) \rightarrow \mathcal{G}.$$

So, when the formal sentence expressing the consistency of PA is assumed, $\mathcal{G}$ becomes provable. This is just what Dummett seems to have in mind when he says that the reasoning can be formalized in the theory.

At this stage perhaps the reader may be misled into thinking that—via the Second Incompleteness Theorem—the truth of $\mathcal{G}$ may be directly derived from that of the formal consistency statement without the need to mention the numerical instances of $\mathcal{G}$. If we have

$$\mathsf{PA} \vdash Cons(\mathsf{PA}) \rightarrow \mathcal{G}$$

[3] PA proves the single numerical instances of $\mathcal{G}$ without proving $\mathcal{G}$ itself. This fact can be easily verified as follows. Suppose to the contrary that $\mathsf{PA} \nvdash \neg\mathsf{Prf}(\overline{n}, \overline{\ulcorner \mathcal{G} \urcorner})$ for a certain $n \in \mathbb{N}$. From Δ_0-decidability of PA, we can infer $\mathsf{PA} \vdash \mathsf{Prf}(\overline{n}, \overline{\ulcorner \mathcal{G} \urcorner})$ and so $\mathcal{G}$ would be provable by means of the proof π whose Gödelian code is n.

then, via the soundness of PA with respect to $\mathcal{N}$, we know that the same implication is also true-in-$\mathcal{N}$, i.e.

$$\mathcal{N} \vDash Cons(\mathsf{PA}) \rightarrow \mathcal{G}.$$

Thus—so the objection goes—why not assume the consistency of PA and then derive the truth of $\mathcal{G}$ by *modus ponens*?[4]

However, the reasoning of those who think of this strategy as a way of avoid numerical instances proceeds too hastily. Suppose we feel confident in Gentzen's consistency proof with transfinite induction up to ϵ_0. Then, the next step should be that of proving the truth-in-$\mathcal{N}$ of the formal consistency statement $Cons(\mathsf{PA})$ by assuming the consistency of PA. As we know from Theorem 13.2.5 and Corollary 13.2.6, both $\mathcal{N} \vDash Cons(\mathsf{PA})$ and $\mathcal{N} \vDash \mathcal{G}$ can be ultimately seen as specific instances of a more general corollary to Σ_1-completeness: *every independent* Π_1 *proposition is true-in-*$\mathcal{N}$. In particular, the two arguments for $\mathcal{N} \vDash Cons(\mathsf{PA})$ and $\mathcal{N} \vDash \mathcal{G}$ produced in this way are isomorphic, i.e. one proof can be turned into the other by uniformly replacing the numerical instances of $Cons(\mathsf{PA})$ with those of $\mathcal{G}$ and vice versa. This is to say that, once one decides to achieve the truth of $\mathcal{G}$ via the Second Incompleteness Theorem, the priority question about $\mathcal{G}$ and its numerical instances boils down to the same question about $Cons(\mathsf{PA})$ and its numerical instances — $\neg\mathsf{Prf}(0, \ulcorner\bot\urcorner), \neg\mathsf{Prf}(1, \ulcorner\bot\urcorner), \neg\mathsf{Prf}(2, \ulcorner\bot\urcorner), \ldots$ — without any logical or epistemological gain.

To resume, the five steps indicated above can be condensed into the following proof.

1.	$\mathcal{N} \vDash \mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$	by contradiction
2.	$\mathsf{PA} \vdash \mathsf{Prf}(\overline{n}, \overline{\ulcorner\mathcal{G}\urcorner})$	Δ_0-completeness
3.	$\mathsf{PA} \vdash \mathcal{G}$	from (1) by def. of $\mathsf{Prf}(x, y)$
4.	$\mathsf{PA} \nvdash \mathcal{G}$	First Incompl. Theorem
5.	$\bot$	from 3, 4
6.	$\mathcal{N} \nvDash \mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$	from 1, 5
7.	$\mathcal{N} \vDash \neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$	Tarski's definition
8.	for all $n \in \mathbb{N}$: $\mathcal{N} \vDash \neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$	arbitrariness of n
9.	$\mathcal{N} \vDash \forall x \neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$	Tarski's definition
10.	$\mathcal{N} \vDash \mathcal{G}$	def. of $\mathcal{G}$.

The alert reader might ask why, in a formal argument supporting the truth of $\mathcal{G}$, even if non-standard, one should resort to the *truth* of the general instance $\neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ instead of directly stressing the *provability* of $\neg\mathsf{Prf}(\overline{n}, \overline{\ulcorner\mathcal{G}\urcorner})$ within PA. This is, indeed, what Tennant does in order to provide a deflationist account of Dummett's argument (Tennant 2002). But this way of reasoning would actually misunderstand the very spirit of Dummett's argument, according to which the knowledge that every $\neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ is true descends from the knowledge that every $\neg\mathsf{Prf}(\overline{n}, \overline{\ulcorner\mathcal{G}\urcorner})$ is provable in PA.

[4] A version of this argument to reach the truth of $\mathcal{G}$ has been proposed in Longo (2011).

Dummett's original line of thought was shored up by Crispin Wright in 1995. Below we formulate Wright's argument in a slightly modified way with respect to Serény's presentation (Serény 2011, p. 51).

1.	$\mathcal{N} \vDash \mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$,	by contradiction
2.	$\mathsf{PA} \vdash \mathsf{Prf}(\overline{n}, \overline{\ulcorner \mathcal{G} \urcorner})$	Δ_0-completeness
3.	$\mathsf{PA} \vdash \mathcal{G}$	from (1) by def. of $\mathsf{Prf}(x, y)$
4.	$\mathsf{PA} \vdash \neg\mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$	diagonalization lemma
5.	$\mathsf{PA} \vdash \forall x \neg \mathsf{Prf}(x, \overline{\ulcorner \mathcal{G} \urcorner})$	definition of Prov(x)
6.	$\mathsf{PA} \vdash \neg\mathsf{Prf}(\overline{n}, \overline{\ulcorner \mathcal{G} \urcorner})$	universal specification
7.	$\mathsf{PA} \vdash \bot$	from (2) and (6)
8.	$\mathcal{N} \nvDash \mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$	by (intuitionistic) refutation of (1)
9.	$\mathcal{N} \vDash \neg\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$	Tarski's definition
10.	$\forall n \in \mathbb{N}\colon \mathcal{N} \vDash \neg\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$	arbitrariness of n
11.	$\mathcal{N} \vDash \forall x \neg\mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$	Tarski's definition
12.	$\mathcal{N} \vDash \neg\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$	definition of $\mathsf{Prov}(x)$
13.	$\mathcal{N} \vDash \mathcal{G}$	diagonalization lemma

It is clear enough that Dummett's and Wright's arguments overlap in sharing the principle of *Universal Generalization* with n as the *generic* element. Nonetheless, Wright's own purpose is to provide grounds for the affirmation of $\mathcal{G}$ *without assuming the consistency of* PA. Which is to say that Wright de-emphasizes consistency in favour of the construction of $\mathcal{G}$ as the fixed point of $\neg\mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$ and the recursive verifiability of the truth of $\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$. But, is the goal of doing without consistency actually accomplished by Wright? Well, this is admittedly a tricky point. On the one hand, we are left with the diagonal construction of $\mathcal{G}$ without comfort in the First Incompleteness Theorem and so in the hypothesis of the consistency of PA. On the other, an unsympathetic reader might note that the rule enabling us to infer $\neg\psi$ whenever ψ leads to a contradiction seems to be a valid principle of reasoning only when considered in a *consistent* system.

However, both the foregoing proofs are not intuitionistically acceptable for the simple fact that the Tarskian biconditional defining the truth of a negated formula in essence delivers bivalence (see Piazza and Pulcini 2013).[5] Nonetheless, these proofs can be turned into fully intuitionistic arguments (as Wright actually does for his own argument) once one makes the two notions of truth and provability coincide. An intuitionistic non-standard proof of $\mathcal{G}$ is reported below. It may improve the reader's grasp to provide it in a natural deduction style, namely:

[5] In PA the principle of bivalence can be presented through the following two clauses: for any formula α,

(1) $\mathcal{N} \vDash \alpha$ or $\mathcal{N} \vDash \neg\alpha$ (either α or its negation is true),
(2) $\mathcal{N} \vDash \alpha \Rightarrow \mathcal{N} \nvDash \neg\alpha$ (there are only two truth-values: *true* and *not-true* (= false)).

Put in these terms, the Tarskian biconditional $\mathcal{N} \vDash \alpha \Leftrightarrow \mathcal{N} \nvDash \neg\alpha$ just expresses bivalence: the rightward direction is given by (2), while the leftward one can be easily restated as (1).

$$
\begin{prooftree}
\AxiomC{$[\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)]$}
\LeftLabel{$\exists\mathcal{I}$}
\UnaryInfC{$\exists x \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$}
\LeftLabel{\textit{Def.}}
\dottedLine
\UnaryInfC{$\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$}
\AxiomC{(\textit{diag. lemma})}
\noLine
\UnaryInfC{$\vdots$}
\noLine
\UnaryInfC{$\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner) \to \neg \mathcal{G}$}
\RightLabel{$\to \mathcal{E}$}
\BinaryInfC{$\neg \mathcal{G}$}
\AxiomC{$\ulcorner \pi \urcorner = n$}
\noLine
\UnaryInfC{$\vdots$}
\noLine
\UnaryInfC{$\mathcal{G}$}
\RightLabel{$\neg \mathcal{E}$}
\BinaryInfC{$\bot$}
\RightLabel{$\neg \mathcal{I}$}
\UnaryInfC{$\neg \mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$}
\RightLabel{$\forall \mathcal{I}$}
\UnaryInfC{$\forall x \neg \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$}
\RightLabel{\textit{Def.}}
\dottedLine
\UnaryInfC{$\mathcal{G}$}
\end{prooftree}
$$

However, it hardly needs to be said that this proof trespasses upon the territory of PA. The explanation is that the proof π extracted by the Gödelian number n does not exist because the natural number n in turn does not actually exist: it is just assumed by contradiction.

An alternative way of accounting for the failure of PA to 'import' this proof refers to the general impossibility of internalising the following conditional[6]

$$\mathsf{PA} \vdash \mathsf{Prov}(\overline{\ulcorner \psi \urcorner}) \Rightarrow \mathsf{PA} \vdash \psi.$$

As Jaroslav Peregrin puts it:

> It is a plain fact that we can demonstrate the truth of $\mathcal{G}$, and hence prove it in the intuitive sense of the word. To be sure, we are not able to do it wholly *within* PA, for the demonstration requires to step out of PA and look at it, as it were, 'from outside'. (Peregrin 2007, p. 7)

Indeed, recall that, according to Löb's theorem,

$$\mathsf{PA} \vdash \mathsf{Prov}(\overline{\ulcorner \psi \urcorner}) \to \psi$$

only in the case that ψ is a theorem of PA. Without the limitations dictated by Löb's theorem, we could write the following proof:

$$
\begin{prooftree}
\AxiomC{$[\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)]$}
\LeftLabel{$\exists\mathcal{I}$}
\UnaryInfC{$\exists x \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$}
\LeftLabel{\textit{Def.}}
\dottedLine
\UnaryInfC{$\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$}
\AxiomC{(\textit{diag. lemma})}
\noLine
\UnaryInfC{$\vdots$}
\noLine
\UnaryInfC{$\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner) \to \neg \mathcal{G}$}
\RightLabel{$\to \mathcal{E}$}
\BinaryInfC{$\neg \mathcal{G}$}
\AxiomC{$[\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)]$}
\LeftLabel{$\exists\mathcal{I}$}
\UnaryInfC{$\exists x \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$}
\LeftLabel{\textit{Def.}}
\dottedLine
\UnaryInfC{$\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$}
\AxiomC{$\vdots$}
\noLine
\UnaryInfC{$\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner) \to \mathcal{G}$}
\RightLabel{$\to \mathcal{E}$}
\BinaryInfC{$\mathcal{G}$}
\RightLabel{$\neg \mathcal{I}$}
\BinaryInfC{$\bot$}
\RightLabel{$\neg \mathcal{I}$}
\UnaryInfC{$\neg \mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$}
\RightLabel{$\forall \mathcal{I}$}
\UnaryInfC{$\forall x \neg \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$}
\RightLabel{\textit{Def.}}
\dottedLine
\UnaryInfC{$\mathcal{G}$}
\end{prooftree}
$$

This proof is obtained from the previous one by replacing the rightmost branch of the former demonstrative tree with a proof explicitly resorting to Löb's Theorem.

[6] The proof of this conditional is as follows. Assume that $\mathsf{PA} \vdash \mathsf{Prov}(\overline{\ulcorner \psi \urcorner})$. Since $\mathsf{Prov}(\overline{\ulcorner \psi \urcorner}) \equiv \exists y \mathsf{Prf}(y, \overline{\ulcorner \psi \urcorner}) \in \Sigma_1$, from Σ_1-completeness we get $\mathcal{N} \vDash \exists y \mathsf{Prf}(y, \ulcorner \psi \urcorner)$, i.e., for a certain $n \in \mathbb{N}$, $\mathcal{N} \vDash \mathsf{Prf}(n, \ulcorner \psi \urcorner)$. This latter means that we can take n and decode it into a PA-proof of ψ. In other words, $\mathsf{PA} \vdash \psi$.

13.4 Is it Possible to Prove $\mathcal{G}$ Bypassing its Numerical Instances?

We turn now to *formal* non-standard arguments. Once we overcome the spell of the naive standard argument, the problem which looms large is whether it is possible to prove $\mathcal{G}$ by ruling out the presence of its numerical instances in some formal version of the standard argument. This is what the proof displayed below takes pains to do (Serény 2011):

1.	$\mathcal{N} \models \mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$	by contradiction
2.	$\mathsf{PA} \vdash \mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$	Σ_1-completeness
3.	$\mathsf{PA} \vdash \mathcal{G}$	from (1) by definition of $\mathsf{Prov}(x)$
4.	$\mathsf{PA} \vdash \neg\mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$	diagonalization lemma
5.	$\mathsf{PA} \vdash \bot$	from (2) and (4)
6.	$\mathcal{N} \not\models \mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$	refutation of (1) by consistency
7.	$\mathcal{N} \models \neg\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$	Tarski's definition
8.	$\mathcal{N} \models \mathcal{G}$	diagonalization lemma

In the above proof only the predicate $\mathsf{Prov}(x)$ officially occurs. However, the proof may be further simplified into a proof in which the predicate $\mathsf{Prov}(x)$ is likewise dispensable:

1.	$\mathsf{PA} \nvdash \neg\mathcal{G}$	First Incompleteness Theorem
2.	$\mathcal{N} \models \neg\mathcal{G} \Rightarrow \mathsf{PA} \vdash \neg\mathcal{G}$	for $\neg\mathcal{G} \in \Sigma_1$ by Σ_1-completeness
3.	$\mathsf{PA} \nvdash \neg\mathcal{G} \Rightarrow \mathcal{N} \not\models \neg\mathcal{G}$	contraposition
4.	$\mathcal{N} \not\models \neg\mathcal{G}$	from (1) and (3) by *Modus Ponens*
5.	$\mathcal{N} \models \mathcal{G}$	Tarski's definition

Nevertheless, we should take care to avoid being led astray by the lack of any reference to the numerical instances of $\mathcal{G}$ in the above proofs. Both these proofs, in the main, are the offshoot of Σ_1-completeness. And Σ_1-completeness cannot be proved without resorting to the correspondence between *truth* and *provability* at the level of the Δ_0-formulas, here given by the numerical instances of φ. One could put it like this: standard proofs do not count as instance-free, much like proofs using a certain lemma requiring the excluded middle do not count as intuitionistic. Summarizing: Σ_1-completeness demands Δ_0-completeness and Δ_0-completeness is reached under Δ_0-decidability: indeed, Σ_1-completeness and Δ_0-decidability are equivalent (Rautenberg 2000). Therefore, the moral we must draw from this is that *any argument designed to operate without the numerical instances of $\mathcal{G}$ actually refers to Σ_1-completeness and so, implicitly, to the numerical instances of $\mathcal{G}$.*

The following proof registers the 'explosion' of the formalized standard argument, so as to make explicit the intervention of Σ_1-completeness.

1.	$\mathcal{N} \models \mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$	by contradiction
$\Rightarrow 1a.$	$\mathsf{PA} \nvdash \exists x \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$	by contradiction
$\Rightarrow 1b.$	$\exists n \in \mathbb{N}\colon \mathcal{N} \models \mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$	from (1) by Tarskian definition
$\Rightarrow 1c.$	$\mathsf{PA} \vdash \mathsf{Prf}(\overline{n}, \overline{\ulcorner \mathcal{G} \urcorner})$	Δ_0-completeness
$\Rightarrow 1d.$	$\mathsf{PA} \vdash \exists x \mathsf{Prf}(x, \overline{\ulcorner \mathcal{G} \urcorner})$	$\exists$-introduction
$\Rightarrow 1e.$	$\mathsf{PA} \vdash \exists x \mathsf{Prf}(x, \overline{\ulcorner \mathcal{G} \urcorner})$	contradiction $(1a)$, $(1d)$
2.	$\mathsf{PA} \vdash \mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$	by def. of $\mathsf{Prov}(x)$
$\Rightarrow 2a.$	$\mathsf{PA} \vdash \exists x \mathsf{Prf}(x, \overline{\ulcorner \mathcal{G} \urcorner})$	by def. of $\mathsf{Prov}(x)$
$\Rightarrow 2b.$	$\mathcal{N} \models \exists x \mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$	soundness of PA w.r.t. $\mathcal{N}$
$\Rightarrow 2c.$	$\exists n \in \mathbb{N}\colon \mathcal{N} \models \mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$	Tarskian definition
3.	$\mathsf{PA} \vdash \mathcal{G}$	by def. of $\mathsf{Prf}(x, y)$
4.	$\mathsf{PA} \vdash \neg\mathsf{Prov}(\overline{\ulcorner \mathcal{G} \urcorner})$	diag. lemma
5.	$\mathsf{PA} \vdash \bot$	from (2) and (4)
6.	$\mathcal{N} \not\models \mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$	refutation of (1) by consistency
7.	$\mathcal{N} \models \neg\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$	Tarski's definition
8.	$\mathcal{N} \models \mathcal{G}$	diagonalization lemma

Pace Serény, then, we have now sufficient reason to conclude that there is no way of reaching the truth of $\mathcal{G}$ without numerical instances playing a part in the process. For Serény, Wright's version is no more than a redundant version of the standard argument when one inserts disguising *detours* involving the predicate $\neg\mathsf{Prf}(x, \ulcorner \mathcal{G} \urcorner)$ and therefore the numerical instances of $\mathcal{G}$. Sefrény also indicates how to prune Wright's version in order to transform it into the standard version: "$\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$ in effect plays the role of $\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$ […] consequently the simplest way to make explicit the logic of the argument is to replace in it every occurrence of $\mathsf{Prf}(n, \ulcorner \mathcal{G} \urcorner)$ with $\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$" (Serény 2011, p. 52).

The tactic Serény employs to dismiss non-standard arguments comes across as trivially sound under the mainstream naive interpretation by which $\mathcal{G}$ declares its own unprovability. Again, the First Incompleteness Theorem kills two birds with one stone by establishing at the same time the independence of $\mathcal{G}$ and the truth of the same sentence. The truth of $\mathcal{G}$ is supposed to follow from the First Incompleteness Theorem *via* the metatheoretical interpretation of $\mathcal{G}$. And if we embrace this interpretation, it will come as no surprise to learn that non-standard arguments collapse into the standard one. But is this interpretation innocent? The standard naive argument improperly transforms Gödel's heuristic argument, at the outset of his 1931 paper, into a proof of the truth of $\mathcal{G}$. In doing so, it must rely on two assumptions: (i) the provability predicate $\mathsf{Prov}(x)$ is actually able to express intensionally the metamathematical property of being a theorem of PA; (ii) the biconditional connective 'if and only if' is capable of faithfully transmitting intensional metamathematical meanings.

$$\mathcal{G} \;\leftrightarrow\; \neg\mathsf{Prov}(\ulcorner \mathcal{G} \urcorner)$$
$\mathcal{G}$ says '$\mathcal{G}$ *is unprovable*'

But why should we treat the biconditional '$\leftrightarrow$' as *says*? If the response is to be compelling, one must also explain why this interpretation is to be preferred to the other equivalent ways of formally expressing the provability predicate on the market (e.g. the Rosser predicate, Arai 1990): why should we agree to interpret the Gödelian provability predicate as intensionally representing the metamathematical notion of provability? The answers to these questions are complicated, but fortunately for our present purposes it is not mandatory to address them.[7] The important thing here is that even if the advocate of the standard view does not presume that the relevant commitments are costless, these costs are however unacceptably high because of unclear or controversial philosophical assumptions.

13.5 Non-standard Proofs as Prototype Proofs

Now let us anticipate an objection. The aficionados of the standard view are likely to protest that in non-standard arguments, as well as in the proof itself of Σ_1-completeness, we avoid mentioning any *individual* numerical instance of the proposition under consideration; rather, we focus on generic instances, i.e. instances obtained from $\forall x \psi(x)$ by replacing the variable x with a *generic* number n which just *takes all individual natural numbers as values*. Thus, appealing to the formula $\psi(n)$—so the objection goes—is tantamount to appealing to a universal quantification. There is a good deal of force in this objection. Evidently, there is more at stake here than the technical question of whether the numerical instances of $\mathcal{G}$ are dispensable for arriving at the truth of $\mathcal{G}$. In other words, the objection has the merit of forcing us explicitly to confront the question of how proofs which incorporate generic elements are to be epistemologically understood.[8]

It has to be said that the presence of generic elements inside a non-standard proof has already been addressed in the ongoing debate. Crispin Wright submits the following:

> Intuitionistically a demonstration of arithmetical '$(\forall x)(Ax)$' is any construction which we can recognise may be used, for an arbitrary natural number k, to accomplish a demonstration of 'Ak'. (Call this principle *Generality*). Well, the reasoning [leading to the truth of $\mathcal{G}$] will evidently go through, if at all, then for an arbitrary choice of 'k'. So it ought to be acknowledged to constitute, by *Generality*, an [intuitionistic]-demonstration of U [that is $\mathcal{G}$]. (Wright 1995b, p. 87)

[7]For the debate see Auerbach (1985), Milne (2007), Heck (2007), Halbach and Visser (2014a, b).

[8]Problems surrounding genericity have always worried logicians. Consider for example the medieval puzzle discussed by Buridan, Burley and Ockham about the man who has promised a horse to someone. For him not to keep his promise, it suffices that he shows the promisee every actual horse, and ask: have I promised you *this* one in particular? So, running through all existing horses, there would be none that he had promised. Ockham's solution is: the proposition 'I promise you a horse' is equivalent to 'I promise you this horse or that horse or that … and so on', but from this does not follow: 'I promise you this horse or I promise you that horse … and so on'. See Klima (2009, pp. 196–199).

Moreover, Michael Detlefsen noticed the same feature in possible arguments supporting the truth of the formal consistency statement $Cons(\mathsf{PA})$ (Detlefsen 1979). Our suggestion, however, is that the issue of genericity goes deeper, since it is fundamental in order to assess the epistemological nature of $\mathcal{G}$. What has not previously been noticed, indeed, is that the notion of genericity is a release from any questioning about the priority between $\mathcal{G}$ and its instances in the strand of thought leading to the truth of $\mathcal{G}$.

Let us elaborate. From a proof-theoretical point of view, proofs which exhibit generic elements can be seen as 'prototypes' applicable to each specific instance of x. The term 'prototype proof' was suggested by Jacques Herbrand in 1931 with reference to the intuitionistic universal quantification: "when we say that a theorem is true for all x, we mean that for each x individually it is possible to iterate its proof, which may just be considered a prototype of each individual proof" (Herbrand 1931, Longo 2000). Hence a prototype proof is executed by focusing on (the properties of) a certain generic element g from the set S the quantifier ranges over, so that the proof exploits only the assumption that g belongs to S.

Prototype proofs are at home in elementary number theory. Consider Euler's proof of Diophantus' proposition:

Proposition 13.5.1 *There is no natural number m of the form* $4n + 3$ *which equals the sum of two integer squared numbers.*

Proof By contradiction we assume the existence of an $m = 4n + 3$ enjoying the property denied in the claim. Since m is *odd*, it should equal the sum of an *odd* squared number with an *even* squared number. Let's say, for a certain pair $k, q \in \mathbb{N}$, $m = (2k)^2 + (2q + 1)^2$, namely $m = 4k^2 + 4q^2 + 4q + 1$ and thence $m = 4(k^2 + q^2 + q) + 1$. $4n + 1$ and $4n + 3$ being mutually exclusive numerical forms, we finally achieve our contradiction. □

This proof manifestly counts as a prototype. Notable about it is the way in which all parameters have a proper degree of generality: the parameter m has the maximal degree, inasmuch as it ranges over the whole set $\mathbb{N}$, the integer n comes as a function of m (a degree of generality just one step below m), and k and q have the same minimal degree of generality, depending on n.

Similarly, non-standard arguments for the truth of $\mathcal{G}$ are instances of prototype arguments insofar as they display a reasoning *uniform* with regard to the generic natural number n inside the formulas $\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ and $\neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$. Inasmuch as prototype proofs express a 'pure' demonstrative method in elementary number theory, non-standard arguments do not stand insulated from the deductive practice of number theory. So, the fact that the truth of $\mathcal{G}$ may be achieved via prototype arguments lends substance to the thought of the arithmetical nature of $\mathcal{G}$.

Let us address a couple of points important to our discussion. The first concerns first-order logic, which, as is well-known, is able to take into account prototypical reasoning. In Gentzen's system of natural deduction a quantification rule authorizes us to infer that all objects have some property from the fact that an arbitrary object has that property:

$$\frac{\alpha(t)}{\forall x \alpha[x/t]} \forall\mathcal{I}.$$

where t must not occur in any assumption on which α depends and $\alpha[x/t]$ is the result of replacing all occurrences of t in α by x. Thus, having insisted that the argument leading to the truth of $\mathcal{G}$ is prototypical, one might wonder why the kind of genericity expressed by first-order logic is not enough to achieve a formal proof of $\mathcal{G}$. The reason is that the universal quantification of $\mathcal{G}$ ranges *exactly* over the set of natural numbers, whereas the 'ontology' of PA—i.e. the set of possible terms that may replace a variable within PA—does not exactly coincide with $\mathbb{N}$. This is the phenomenon known as the non-categoricity of PA and it is a model-theoretic byproduct of incompleteness.

The second point to emphasize is epistemological. We are arguing here that the specific prototypical nature of the argument for the truth of $\mathcal{G}$ is what confers to $\mathcal{G}$ its very epistemological status. The assumption at work is that the resort to prototype reasoning is actually *unavoidable* in order to achieve the truth of $\mathcal{G}$. In particular, this last claim amounts to affirming that there is no way to 'decide' $\mathcal{G}$ by resorting to mathematical induction. This last fact is undisputed: if $\mathcal{G}$ were provable by ordinary induction on $\mathbb{N}$, then this proof would be formalizable *within* PA, which is clearly not the case.

Recognizing the prototypical nature of non-standard arguments goes hand in hand with removing the epistemological relation of priority from consideration. The question is this: do prototype proofs really need to pass through the numerical instances of the proved statement $\forall x \psi(x)$? It seems fair to answer in the negative. But neither can we read this answer as implying that $\forall x \psi(x)$ is epistemologically prior to its numerical instances. Unlike the inductive ones, indeed, prototype proofs by no means require that we talk about *individual* instances of a formula $\forall x \psi(x)$—i.e. instances produced by substituting x with naturals in the flesh: $\psi(0), \psi(1), \psi(2), \ldots$. They require instead that we are able to talk about what we might call, in the absence of a well-established terminology, the *universal instances* of $\forall x \psi(x)$, i.e. instances generated by replacing the variable x with a generic object. So, it should be said straightaway that such objects have a remarkable hybrid character, in that they represent something midway between universal quantification and individual numerical instances. This hybrid character also constitutes, it seems to us, the most fundamental obstacle to their characterization. On the one hand, prototype proofs work by manipulating, logically or algebraically, formulas which stand for singular objects; on the other, these singular objects retain a dimension of generality throughout the manipulations imposed by the proof. It is from such a dialectical interplay between universal quantification and numerical instances that prototype proofs draw their persuasive power.

It is instructive, then, to reconsider how the unpacked standard proof in Sect. 13.3 takes us down the road of $\mathcal{G}$'s truth: $\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ and its negation $\neg\mathsf{Prf}(n, \ulcorner\mathcal{G}\urcorner)$ play the role of the instances in which the universal and the particular merge. On the one hand, one needs to shift from $\mathcal{G}$ to its (universal) instances so as to reduce the logical complexity and trigger Σ_1-completeness by means of the Δ_0-completeness.

On the other, one must take care of the generality of n so as to go back legitimately from the (universal) instances to the universal quantification and thence get $\mathcal{G}$. As remarked in Sect. 13.2, Σ_1-completeness is proved *via* Δ_0-completeness, which, in turn, presupposes Δ_0-decidability. In other words, PA is able to 'recognise deductively' true Σ_1-formulas, i.e. formulas expressible as $\exists x\psi(x)$ where any single instance $\psi(\overline{0}), \psi(\overline{1}), \ldots$ belongs to the class of the formulas Δ_0.

The reason why Σ_1-completeness holds true can be fully grasped once this problem is algorithmically addressed. Indeed, the recursive algorithm for the verification of $\exists x\psi(x)$ can be thought of as follows:

(i) by Δ_0-decidability, PA is able to check, progressively, each one of the numerical instances $\psi(\overline{0}), \psi(\overline{1}), \ldots$,
(ii) once this procedure encounters an $n \in \mathbb{N}$ such that $\psi(\overline{n})$ is true, PA is able to prove it (Δ_0-completeness) and so to infer $\mathsf{PA} \vdash \exists x\psi(x)$.

Now should emerge the inescapable role played by the numerical instances of arithmetical formulas in the algorithmic mechanism underlying the property of Σ_1-completeness.

Most importantly, we claim that the effectiveness of prototype proofs as argumentative tools is closely tied to an epistemological balance between the truth of the universal quantification and that of its numerical instances. In other words, it is the very activation of universal instances that produces the epistemological balance at issue. This balance wards off an equivocation: that a specific priority can be established between the truth of the universally quantified like $\mathcal{G}$ and that of its numerical instances.

13.6 Concluding Thoughts

Our fundamental concern in this paper has been the problem of how to articulate our understanding of the truth of the independent Gödel sentence $\mathcal{G}$. We have argued that:

1. Once the rhetorical clouds of the naive standard argument have been dispersed, the *formal* renditions of this argument need to introduce the numerical instances of $\mathcal{G}$ insofar as the Σ_1-completeness property is involved. Therefore, the idea of providing grounds for affirming $\mathcal{G}$ without mentioning its numerical instances turns out to be misguided.
   First conclusion: standard arguments are non-standard arguments in a hurry: some deductive passages are omitted but implicitly assumed.
2. The Second Incompleteness Theorem provides a parallel path to the truth of $\mathcal{G}$ by suggesting that one can derive $\mathcal{G}$ from $Cons(\mathsf{PA}) \rightarrow \mathcal{G}$ by *modus ponens*. But this strategy is ineffective in getting rid of numerical instances.

Second conclusion: in the proof leading to the truth of $Cons(\mathsf{PA})$, we are likewise forced to refer to the numerical instances of $Cons(\mathsf{PA})$.

3. Arguments leading to the truth of $\mathcal{G}$ can be slotted into the category of prototype arguments to the extent that they need to stress the generic natural number n. Thus, arguments establishing the truth of $\mathcal{G}$ cannot do without mentioning its numerical instances, and these instances are of a special kind: they are produced by instantiating $\mathcal{G}$ with the generic natural number n. We called these specific mathematical objects *universal instances*.
   Third conclusion: in the case of $\mathcal{G}$ the debate about the priority between universal quantification and its numerical instances has so far proceeded from a false alternative: either we move from $\mathcal{G}$ to the instances or from the instances to $\mathcal{G}$. Such an alternative arises from a misconception of the very nature of the proof of the truth of $\mathcal{G}$. Once one realises the prototypical nature of the arguments leading to the truth of $\mathcal{G}$, the dilemma vanishes.
   Fourth conclusion: the reason we are inclined to regard $\mathcal{G}$ as stating an arithmetical property is that prototype arguments belong to the perfectly ordinary practice of number theory. Thus, in the absence of any convincing arguments for the genuine metatheoretical nature of $\mathcal{G}$, the following two theses should be rejected:

   (a) $\mathcal{G}$ cannot express a clear arithmetical meaning independently of any kind of metatheoretical commitment.
   (b) (*Isaacson thesis*) If we are to give a proof for any true sentence (expressed in the language of PA) which is independent of PA, then we will need to appeal to ideas that go beyond those that are required in understanding PA (Isaacson 1987; Smith 2008).

Acknowledgments We would like to thank two anonymous readers and Giuseppe Longo for comments and suggestions. G.P. acknowledges the support from FAPESP Post-Doc Grant 2013/22371-0, São Paulo State, Brazil.

References

Arai, T. (1990). Derivability conditions on Rosser's provability predicates. *Notre Dame Journal of Formal Logic*, *31*, 487–497.
Auerbach, D. (1985). Intentionality and the Gödel theorems. *Philosophical Studies*, *48*, 337–351.
Detlefsen, M. (1979). On interpreting Gödel's second theorem. *Journal of Philosophical Logic* 8(1), 297–313.
Dummett, M. (1963). The philosophical significance of Gödel's theorem. *Ratio, 5*, 140–155. Reprinted in *Truth and other enigmas* (1978). London: Duckworth.
Gaifman, H. (2000). What Gödel's incompleteness result does and does not show. *The Journal of Philosophy*, *97*, 462–470.
Gentzen, G. (1969). In M. E. Szabo (Eds.), *The collected papers of Gerhard Gentzen*. North-Holland, Amsterdam.
Halbach, V., & Visser, A. (2014a). Self-reference in arithmetic I. *The Review of Symbolic Logic*, *7*, 671–691.

Halbach, V., & Visser, A. (2014b). Self-reference in arithmetic II. *The Review of Symbolic Logic*, *7*, 692–712.

Heck, R. G. (2007). Self-reference and the languages of arithmetic. *Philosophia Mathematica*, *15*, 1–29.

Herbrand, J. (1931). Sur la non-contradiction de l'Arithmétique. *Journal für die reine und angewandte Mathematik*, *166*, 1–8.

Isaacson, D. (1987). Arithmetical truth and hidden higher-order concepts. In The Paris Logic Group (Ed.), *Logic Colloquium '85* (pp. 147–169). Amsterdam: North-Holland.

Klima, G. (2009). *John Buridan*. Oxford: Oxford University Press.

Longo, G. (2000). Prototype proofs in type theory. *Mathematical Logic Quarterly*, *46*(3), 257–266.

Longo, G. (2011). Reflections on (concrete) incompleteness. *Philosophia Mathematica*, *19*(3), 255–280.

Milne, P. (2007). On Gödel sentences and what they say. *Philosophia Mathematica*, *15*, 193–226.

Peregrin, J. (2007). Gödel, truth & proof. *Journal of Physics: Conference Series 82*, 1–10.

Piazza, M., & Pulcini, G. (2013). Strange case of Dr. Soundness and Mr. Consistency. In *The logica yearbook 2013* (pp. 161–172). College Publications.

Piazza, M., & Pulcini, G. (2015). A deflationary account of the truth of the Gödel sentence $\mathcal{G}$. In G. Lolli et al. (Eds.), *From logic to practice* (pp. 71–90). Boston studies in the philosophy and history of science. Berlin: Springer.

Rautenberg, W. (2000). *A concise introduction to mathematical logic.* Berlin: Springer.

Serény, G. (2011). How do we know that the Gödel sentence of a consistent theory is true? *Philosophia Mathematica*, *III*, 47–73.

Smith, P. (2008). Ancestral arithmetic and Isaacson's thesis. *Analysis*, *68*, 1–10.

Smoryńsky, C. (1977). *The incompleteness theorems. Handbook of mathematical logic*. North Holland: Amsterdam.

Tennant, N. (2002). Deflationism and the Gödel phenomena. *Mind*, *111*, 551–582.

Wright, C. (1995a). About "The philosophical significance of Gödel's Theorem": Some issues. In B. McGuinness & G. Oliveri (Eds.), *The philosophy of Michael Dummett*. Oxford: Blackwell.

Wright, C. (1995b). Intuitionists are not (Turing) machines. *Philosophia Mathematica*, *3*(3), 86–102.

Chapter 14
More on Systems of Truth and Predicative Comprehension

Carlo Nicolai

Abstract In the paper we survey the known connections between theories that extend a common base theory with typed truth axioms on the one hand and predicative set-existence assumptions on the other. How general can the mutual reductions between truth and comprehension be taken to be? In trying to address this question, we consider (typed) classical, positive truth and predicative comprehension as operations on theories.

Keywords Predicative comprehension · Axiomatic theories of truth · Relative interpretability

14.1 Introduction

The proof-theoretic analysis of axiomatic truth theories has uncovered several connections holding between systems extending a base theory with set-existence axioms and systems extending the same base theory with axioms characterizing a primitive truth predicate. Taken at face value, these results suggest that assumptions on the existence of certain sets can be replaced by suitable semantic assumptions and vice versa. As we shall see shortly, for instance, the extension of Peano Arithmetic with specific assumptions on arithmetically definable sets is inter-reducible with a full, Tarskian truth theory over PA.

C. Nicolai (✉)
Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft,
Munich Center for Mathematical Philosophy, Geschwister-Scholl-Platz 1,
D-80539 München, Germany
e-mail: carlo.nicolai6@gmail.com

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_14

These connections have played an important role in the foundation of mathematics (e.g. in the analysis of the limits of predicativity Feferman 1991), in the analysis of different solutions to the semantic paradoxes (proving more truth-theoretic iterations has been generally considered a virtue of a theory of truth), in the debate concerning the nature of truth (is truth a light or a substantial property?). They have been also suggested as a tool to carry out ontological reductions.[1]

At any rate, these mutual reductions between truth and set existence axioms involve theories built on a fixed base theory, usually Peano Arithmetic. In this work we set the basis for a different approach: truth and predicative comprehension will be taken as functors applying to arbitrary object theories satisfying some minimal requirements. Crucially, the results of applying these functors to the base theory will turn out to be equivalent tools for uncovering the first step of our commitment implicit in the acceptance of the base theory. Furthermore, the generality given by the proposed approach will also be reflected in a more uniform correspondence between set-theoretic and semantic assumptions, as it will be exemplified in Sect. 14.5.

The paper has two main parts. In the first, we survey several results connecting extensions Peano Arithmetic with truth axioms or predicative comprehension axioms. In the second, we introduce the more general setting just announced.

As this is more congenial to the original, second part of the work we will mostly focus, in the first part, on systems of *typed truth*—that is, theories formulated in languages in which the truth predicate applies to sentences not containing the truth predicate itself—and subsystems of ACA. The core argument, which gives rise to several variations, is the folklore mutual reduction between the system of Tarskian, compositional truth CT and ACA itself. We will also take into account the system PT of positive, typed truth, which is a notational variant of CT: its subsystems are not so easily comparable with their Tarskian counterparts but still enjoy several reductions to subsystems of ACA. All of this will be surveyed in Sect. 14.3.

The second part, starting with Sect. 14.4, will investigate a generalization of the systems considered in Sect. 14.3. We define three main operations on recursively enumerable (RE) theories: (i) $\mathsf{T}[\cdot]$ results in a Tarskian truth theory; (ii) $\mathsf{Tp}[\cdot]$ in a typed theory of truth simulating positive inductive definitions; (iii) $\mathsf{PC}[\cdot]$ adds predicative comprehension to the object theory. In order to study $\mathsf{PC}[\cdot]$ and relate it in full generality to the truth theories also studied, a variant of it, $\mathsf{PCS}[\cdot]$, has to be taken into account: it applies to theories axiomatized by schemata in which schematic variables are replaced by 'second-order' variables. The upshot of Sect. 14.4 will be that, modulo mutual interpretability, the extension of arbitrary RE base theories via these three operations yields equivalent results. In Sect. 14.5 we reflect on the significance of the technical work carried out in the previous sections and consider possible extensions.

[1]Cf. Halbach (2014, Chap. 23) and Sect. 14.5 of the present paper.

14.2 Preliminaries

14.2.1 *Theories and Arithmetization*

Unless otherwise specified, arbitrary theories will be formulated in many-sorted, first-order logic. We assume they have a Δ_1^b specification[2]—provably in a weak syntax theory to be defined—and that they are formulated in a Hilbert-style calculus in with Modus Ponens as its only rule of inference.[3]

On occasion, we consider *sequential* theories, namely theories that have a nice coding of sequences, which is *in turn a solid basis* to formalize syntax.[4] We employ $T, U, V, W, \ldots$ to range over arbitrary theories; $A, B, C, \ldots$ are taken to range over finitely axiomatized theories.

Whereas in Sect. 14.3 we work with PA as base theory, so all the usual representations of syntactic notions and operations can be assumed, in Sect. 14.4 we work in the theory S_2^1. S_2^1 is a theory introduced by Sam Buss in Buss (1986) to study polynomial time computability; the functions that are provably recursive in S_2^1 are exactly the p-time computable functions. Remarkably, S_2^1 is finitely axiomatizable and mutually interpretable with the induction-free fragment of first-order arithmetic Q. For details concerning coding in S_2^1, we refer to Buss (1986, 1998). In particular, for any given language $\mathcal{L}_W$, we assume Δ_1^b-definitions $\mathsf{Term}_{\mathcal{L}_W}(x)$, $\mathsf{CTerm}_{\mathcal{L}_W}(x)$, $\mathsf{Fml}_{\mathcal{L}_W}(x)$, $\mathsf{Sent}_{\mathcal{L}_W}(x)$, $\mathsf{Prf}_{\mathcal{L}_W}(x)$, $\mathsf{Proof}_{\mathcal{L}_W}(x, y)$ of the sets of terms, closed terms, formulas, sentences of the language of W, of proofs in the theory W and of the relation of being a proof in W of the $\mathcal{L}_W$-formula y. Also, we take the set of theorems of W to be defined by the $\exists\Delta_1^b$-formula $\mathsf{Pr}_W(x) :\leftrightarrow \exists y\, \mathsf{Proof}_W(x, y)$.[5] As to notational conventions, we will often abuse of Gödel corners and interchangeably employ them and Feferman's dot convention.

---

[2]That is, their axiom set is specified by a formula provably equivalent to formulas containing sharply bounded quantifiers (cf. Hájek and Pudlák 1993, V.4. Definition 4.2) and one bounded existential or universal quantifier only.

[3]Cf. Enderton (2001) for an example of such axioms system for first-order logic.

[4]More precisely, T is sequential iff it directly interprets (i.e. identity is mapped into identity and quantifiers are not relativized) *Adjunctive Set Theory*, that is a theory in a (first-order) language with $\in$ and $=$ whose axioms are:

$$\exists x\, \forall y\; y \notin x \tag{AS1}$$

$$\forall u, v\, \exists x\, \forall y\; (y \in x \leftrightarrow (y \in u \vee y = v)) \tag{AS2}$$

Remarkably, Q is not sequential, but many sequential theories are interpretable in it.

[5]The class of $\exists\Delta_1^b$ formulas is obtained from the class of Δ_1^b-formulas by closing it under unbounded existential quantification.

14.2.2 Reductions

The preferred means of reduction throughout the paper will be *many sorted—possibly non direct—relative interpretability.*

Let T and W be theories containing S^1_2.[6] A *relative translation* of $\mathcal{L}_T$ into $\mathcal{L}_W$ can be described as a pair (δ, F) where δ is a $\mathcal{L}_W$-formula with one free variable—the domain of the translation—and F is a (finite) mapping that takes n-ary relation symbols of $\mathcal{L}_T$ and gives back formulas of $\mathcal{L}_W$ with n free variables.[7] The translation extends to the mapping τ:

- $(R)^\tau(x_1, \ldots, x_n) :\leftrightarrow F(R)(x_1, \ldots, x_n)$;
- τ commutes with the propositional connectives;
- $(\forall x\ \varphi(x))^\tau :\leftrightarrow \forall x\ (\delta(x) \rightarrow \varphi^\tau)$ and $(\exists x\ \varphi(x))^\tau :\leftrightarrow \exists x\ (\delta(x) \wedge \varphi^\tau)$.

An *interpretation* K is then specified by a triple (T, τ, W) such that for all sentences φ of $\mathcal{L}_T$,

$$T \vdash \varphi \Rightarrow W \vdash \varphi^\tau$$

W *locally* interprets T if and only if every finite subsystem of T is interpretable in W. An interpretation is *direct* if and only if it maps identity to identity and it does not relativize quantifiers. We will often not distinguish between an interpretation and the relative translation that supports it.

On occasion, we will employ the notion of *truth definability*, extensively studied by Fujimoto (2010). Truth definability will be mainly used in the first part of the paper, as it assumes a fixed base theory (at least in its original formulation in Fujimoto 2010). Let U, V be theories extending a syntactic base theory B: it is assumed, moreover, that U, but not necessarily also V, results from the extension of B via truth axioms. U is *relatively truth definable* in V if there is a $\mathcal{L}_B$-conservative relative interpretation of U into V. In other words, a truth definition of U in V is a relative interpretation τ_0 that behaves like the identity mapping when applied to truth-free formulas of $\mathcal{L}_U$.

14.2.3 Definable Cuts and Incompleteness

Cuts are initial segments of a *theory's* natural numbers. We will in particular be interested in initial segments definable in theories T containing S^1_2. More precisely, a formula $\varphi(x)$ is called *inductive* in T containing S^1_2 if and only if

$$T \vdash \varphi(0) \wedge \forall y\ (\varphi(y) \rightarrow \varphi(Sy))$$

[6]The relation of containment can be understood as subtheory relation or relative interpretability relation.

[7]The definition generalizes naturally to the many-sorted case by considering multiple domains.

$\varphi(x)$ is a *T-cut* if and only if, additionally, it defines an initial segment of the T-numbers. In other words,

$$T \vdash \forall x, y\ (\varphi(x) \wedge y \leq x \rightarrow \varphi(y))$$

Here it is important to notice that $0, S, \leq$ are understood in T via the assumed interpretation of S^1_2 in T. By a well-know result of Solovay,[8] every inductive formula has a subcut:

Lemma 1 *Let $\varphi(x)$ be inductive in T containing S^1_2. Then there exists a T-cut $\psi(x)$ such that*

$$T \vdash \psi(x) \rightarrow \varphi(x).$$

The ψ in Lemma 1 is obtained by closing $\varphi(\overline{n})$ for each n under transitivity of $\leq$ so that this holds for all $m \leq n$ as well. It is often useful, however, to employ a slightly modified notion of definable cut. The modifications are justified by the following Lemma:

Lemma 2 *Let T interpret S^1_2 and $\mathcal{I}$ be a T-cut. Then we can find a subcut $\mathcal{J}$ of $\mathcal{I}$ such that T proves the following:*

$$\mathcal{J}(x) \wedge \mathcal{J}(y) \rightarrow \mathcal{J}(x + y) \tag{14.1}$$

$$\mathcal{J}(x) \wedge \mathcal{J}(y) \rightarrow \mathcal{J}(x \times y) \tag{14.2}$$

$$\mathcal{J}(x) \wedge \mathcal{J}(y) \rightarrow \mathcal{J}(x\#y). \tag{14.3}$$

where the smash function # is such that $x\#y = 2^{|x| \times |y|}$, and $|x| = \lceil \log_2(x + 1) \rceil$—that is, the upper integer part of the binary logarithm of $x + 1$.

Therefore, in what follows, a definable cut can always be taken to be closed under addition, multiplication and the smash function. Lemma 2 is particularly important as it guarantees that, given a T-cut $\mathcal{I}$—with T again containing S^1_2, we can always restrict our attention to a subcut $\mathcal{J}$ of $\mathcal{I}$ satisfying the axioms of S^1_2. This will be extensively employed in Sect. 14.4: having S^1_2 available on a cut enables one to have a meaningful and smooth arithmetization of the syntax in the cut. This will prove to be crucial in many of the interpretations defined later on.

We will extensively make use of the following:

Lemma 3 (Wilkie, Nelson) Q *interprets* S^1_2 *on a definable cut.*

Q does not have any induction: it cannot thus prove the usual properties of $\leq$. An important part of the proof of Lemma 3 consists in showing how the interpretation is indeed defined on a initial segment of the Q-numbers. This makes possible the downwards preservation of Π_1-sentences (cf. Hájek and Pudlák 1993, V.5(c)) so that, for instance, $\mathsf{S}^1_2 + \mathsf{Con}(U)$ is interpretable in $\mathsf{Q} + \mathsf{Con}(U)$.

[8] Although contained in an unpublished note (cf. Hájek and Pudlák 1993).

Another remarkable fact concerning S^1_2 is that Gödel's Second Incompleteness theorems can be meaningfully stated and proved in it. This makes possible a generalization of the Gödel's second incompleteness theorem, due to Pavel Pudlák in its original form. The beautiful version that we present is due to Albert Visser.

Lemma 4 *Let U be given by a Σ_1-formula. Then U does not interpret* $\mathsf{S}^1_2 + \mathsf{Con}(U)$.

Proof Assume it does. There must be, therefore, a finite subsystem A of U that interprets $\mathsf{S}^1_2 + \mathsf{Con}(U)$. By Σ_1-completeness, S^1_2 proves the formalization of the fact that $A \subset U$. Thus A interprets $\mathsf{S}^1_2 + \mathsf{Con}(A)$. Since A and $\mathsf{S}^1_2 + \mathsf{Con}(A)$ are finite, we have

$$\mathsf{S}^1_2 \vdash \mathsf{Con}(A) \to \mathsf{Con}(\mathsf{S}^1_2 + \mathsf{Con}(A)).$$ □

14.3 Arithmetical Comprehension

In this section we survey the mutual reductions between truth axioms and predicative comprehension in the standard setting, that is when PA is taken as object theory. After a few useful definitions and lemmata, in Sect. 14.3.1 we consider how Tarskian and positive truth can define the comprehension schema of ACA and variants thereof. We then consider the converse direction in Sect. 14.3.2. Section 14.3.3 hints at how to compare iterations of Tarskian truth and predicative comprehension to each other and to systems of type-free truth.

The language $\mathcal{L}_2$ extends the language $\mathcal{L} := \{0, S, +, \times, \leq\}$ of arithmetic with an additional sort for 'sets of natural numbers', or 'reals'.

Definition 1 (i) ACA is formulated in $\mathcal{L}_2$. It extends PA with

$$\exists Y \forall u (u \in X \leftrightarrow \varphi(u)) \tag{CA}$$

where φ does not contain bound set-variables (it may contain set parameters); and

$$\varphi(0) \wedge \forall y(\varphi(y) \to \varphi(Sy)) \to \forall y\, \varphi(y) \tag{S-Ind^2}$$

for $\varphi \in \mathcal{L}_2$.

(ii) To obtain ACA_0 one replaces (S-Ind^2) with the single $\mathcal{L}_2$-sentence

$$0 \in X \wedge \forall y\, (y \in X \to Sy \in X) \to \forall y\ y \in X \tag{A-Ind^2}$$

Lemma 5 ACA_0 *is not interpretable in* PA.

Proof If ACA_0 were interpretable in PA, then ACA_0 would be interpretable in a suitable finite subsystem A of PA. Thus PA would be interpretable in A. But PA is reflexive—that is, it proves the consistency of all its finite subtheories. Therefore,

$$\begin{aligned}&\mathsf{PA} \vdash \mathsf{Con}(A) \rightarrow \mathsf{Con}(\mathsf{ACA}_0)\\&\mathsf{PA} \vdash \mathsf{Con}(\mathsf{ACA}_0)\end{aligned}$$

But $\mathsf{PA} \subset \mathsf{ACA}_0$, so the last line contradicts Gödel's second incompleteness theorem. □

14.3.1 Arithmetical Comprehension from Truth

Definition 2 (CT)

(i) CT extends PAT (i.e. PA in $\mathcal{L} \cup \{\mathsf{Tr}\}$) with the universal closure of

$$\mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \rightarrow (\mathsf{Tr}\, x \underset{.}{\circ} y \;\leftrightarrow\; \mathsf{val}(x) \circ \mathsf{val}(y)) \tag{CT1}$$

$$\mathsf{Sent}_{\mathcal{L}}(x) \rightarrow \;(\mathsf{Tr} \underset{.}{\neg} x \leftrightarrow \neg \mathsf{Tr}\, x) \tag{CT2}$$

$$\mathsf{Sent}_{\mathcal{L}}(x \underset{.}{\wedge} y) \rightarrow \;(\mathsf{Tr}(x \underset{.}{\wedge} y) \;\leftrightarrow\; \mathsf{Tr}\, x \wedge \mathsf{Tr}\, y) \tag{CT3}$$

$$\mathsf{Sent}_{\mathcal{L}}(\underset{.}{\forall} v x) \rightarrow \;(\mathsf{Tr} \underset{.}{\forall} v x \;\leftrightarrow\; \forall y (\mathsf{CTerm}(y) \rightarrow \mathsf{Tr}\, \mathsf{sub}(x, v, y))) \tag{CT4}$$

In (CT1), $\circ \in \{=, \leq\}$. In (CT4), v *codes* a variable.

(ii) $\mathsf{CT}\!\upharpoonright$ is obtained by allowing only $\mathcal{L}$-formulas as instances of the induction scheme of CT;

(iii) $\mathsf{CT}^{-} :=$ (CT1)−(CT4) plus:

$$\mathsf{tot}(x) \rightarrow \mathsf{Tr} x(\ulcorner 0 \urcorner) \wedge \forall y\, (\mathsf{Tr}\, x(\dot{y}) \rightarrow \mathsf{Tr}\, x(\mathsf{S}\dot{y})) \rightarrow \forall y\, \mathsf{Tr}\, x(\dot{y}) \tag{IInd}$$

where

$$\mathsf{tot}(x) :\leftrightarrow \mathsf{Fml}^{1}_{\mathcal{L}}(x) \rightarrow \forall y\, (\mathsf{Tr}\, x(\dot{y}) \vee \mathsf{Tr} \underset{.}{\neg}\, x\, (\dot{y})) \tag{tot}$$

and $\mathsf{Fml}^{1}_{\mathcal{L}}(x)$ expresses that x is a formula of $\mathcal{L}$ with only the first variable free.

CT^{-} is considered in Fischer (2009). It is finitely axiomatized and thus weaker than CT: the latter is in fact reflexive.

We summarize some useful facts about CT and its restricted variants. Most of them, in slightly different form, can also be found for instance in Cantini (1989), Feferman (1991), Halbach (2014).

Lemma 6 *(i) for all* $\varphi(v) \in \mathcal{L}$, $\mathsf{CT}\!\upharpoonright \vdash \forall x\, (\mathsf{CTerm}_{\mathcal{L}}(x) \rightarrow (\mathsf{Tr}\, \mathsf{sub}(\ulcorner \varphi \urcorner, v, x) \leftrightarrow \varphi(\mathsf{val}(x))))$;

(ii) $\forall x\, (\mathsf{Sent}_{\mathcal{L}}(x) \wedge \mathsf{Pr}_{\mathsf{PA}}(x) \rightarrow \mathsf{Tr}\, x)$, *where* $\mathsf{Pr}_{\mathsf{PA}}(\cdot)$ *expresses canonical provability in* PA*;*

(iii) $\mathsf{CT}\!\upharpoonright \vdash \forall x (\mathsf{Fml}^{1}_{\mathcal{L}}(x) \rightarrow \mathsf{tot}(x))$

(iv) $\mathsf{CT}^{-} \vdash \mathsf{Con}(\mathsf{PA})$

(v) $\mathsf{CT} \vdash \mathsf{Sent}_{\mathcal{L}}(\dot{\forall} vz) \wedge \mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \rightarrow (\mathsf{val}(x) = \mathsf{val}(y) \ \rightarrow \ (\mathsf{Tr}(\mathsf{sub}(z, v, x) \ \leftrightarrow \ \mathsf{Tr}(\mathsf{sub}(z, v, y)))))$

Proof (i) is obtained by external induction on φ. For (ii) one employs the induction axioms of CT. (iii) follows from CT2. (iv) is obtained by combining (IInd) and (i) to mimic in CT^- the CT-proof of (ii). (v) also follows from a formal induction on the complexity of the formula z. □

By extending PAT with the schema of Lemma 6(i) taken as axiom, we obtain the theory UTB. In $\mathsf{UTB}\!\restriction$, like in $\mathsf{CT}\!\restriction$, the truth predicate is not allowed into instances of the induction scheme. The next lemma summarizes some well-known facts concerning the interpretability of typed truth in PA.

Lemma 7 *(i)* PA *locally interprets* $\mathsf{TB}\!\restriction$, $\mathsf{UTB}\!\restriction$, TB, UTB*;*
(ii) PA *interprets* $\mathsf{TB}\!\restriction$, $\mathsf{UTB}\!\restriction$, TB, UTB*;*
(iii) PA *interprets* $\mathsf{CT}\!\restriction$*;*
(iv) PA *does not interpret* CT^-, CT.

Proof Ad (i), in a finite subsystem S of UTB instances of the uniform disquotation scheme involve sentences $\varphi \in \mathcal{L}$ such that $\mathsf{lcx}(\varphi) \leq n$ for a standard n, where $\mathsf{lcx}(x)$ a primitive recursive function that, applied to a $\mathcal{L}$-formula φ, yields the number of its logical symbols. We interpret the truth predicate of S with the partial truth predicate, definable in PA, for sentences of complexity up to n. (ii) follows from (i) and Orey's compactness theorem, according to which if, W is locally interpretable in T and T is reflexive, then T interprets W. The proof of (iii) contained in Fischer (2009) works if, instead of the cut-elimination argument of Halbach (1999),[9] one employs the proof of the conservativity of $\mathsf{CT}\!\restriction$ over PA contained in Leigh (2014). A new, direct proof is contained in Enayat and Visser (2014, Theorem 5.1). (iv) is straightforward from Lemma 6(iv). □

Proposition 1 CT *interprets* ACA. *Moreover, the interpretation behaves like the identity mapping on the arithmetical vocabulary.*

Proof A full proof is contained in Halbach (2014): we present some of its details as they will be useful for later proofs.

We define a substitution function for elements of $\mathsf{Fml}^1_{\mathcal{L}}$:

$$\mathsf{sb}(\ulcorner \varphi(v_1) \urcorner, \ulcorner t \urcorner) := \ulcorner \varphi(t) \urcorner$$

We let $\mathsf{sb}(x, y)$ take a default value, e.g. $\ulcorner 0 = 0 \urcorner$, when it is not applied to an element of $\mathsf{Fml}^1_{\mathcal{L}}$ and to a term. The translation ι is then crucially defined by the following clauses:

[9] It has been shown to contain a mistake by Kentaro Fujimoto.

$$(x \in y)^\iota :\leftrightarrow \mathsf{Tr}\,(\mathsf{sb}(x, y)) \qquad y \text{ is satisfied by } x$$

ι maps the arithmetical nonlogical vocabulary of $\mathcal{L}_2$ into itself;

ι commutes with propositional quantifiers;

$$(\forall X\, \varphi)^\iota :\leftrightarrow \forall X^\iota\, (\mathsf{Fml}^1_{\mathcal{L}}(X^\iota) \to (\varphi)^\iota)$$

Set variables and number variables are kept distinct to avoid clashes: this is emphasized by denoting with X^ι the $\mathcal{L}$-variable resulting from the translation. Strictly speaking, X^ι is a variable of the only available sort.

To verify that ι behaves as required, we consider the crucial case of comprehension axioms, i.e. we prove

$$(\exists X\, \forall x\, (x \in X \leftrightarrow \varphi(x))^\iota \tag{14.4}$$

Notice that $\varphi(x)$ *may contain first- and second-order parameters.* Then, in CT:

$$\forall x(\varphi^\iota(x, \mathsf{Tr}(\mathsf{sb}(w, X^\iota))) \leftrightarrow \varphi^\iota(x, \mathsf{Tr}(\mathsf{sb}(w, X^\iota)))) \qquad \text{where } w \text{ is a closed term} \tag{14.5}$$

$$\forall x(\mathsf{Tr}\ulcorner \varphi^\iota(\dot{x}, \mathsf{sb}(w, X^\iota))\urcorner \leftrightarrow \varphi^\iota(x, \mathsf{Tr}(\mathsf{sb}(w, X^\iota)))) \qquad \text{Lemma 6(i) and CT1-4} \tag{14.6}$$

$$\exists Y^\iota\, \forall x(\mathsf{Tr}(\mathsf{sb}(\dot{x}, Y^\iota)) \leftrightarrow \varphi^\iota(x, \mathsf{Tr}(\mathsf{sb}(w, X^\iota)))) \tag{14.7}$$

Essentially, the last line results from the application of the compositional axioms of CT to move the occurrences of the truth predicate in $\varphi^\iota(x, \mathsf{Tsb}(w, X^\iota))$ in front of the formula. In line (14.6), we reason for an arbitrary φ: strictly speaking, an external induction on the complexity of the formula is needed in which the different compositional axioms are employed. The last line follows from the properties of substitution. □

In the proof of Lemma 1, the compositional axioms CT1-4 are only employed in moving from (14.5) to (14.6) and, in particular, to deal with parameters occurring into the comprehension scheme. We then let $\mathsf{ACA}^{\mathsf{pf}}$ to be exactly as ACA but with no parameters allowed to appear into the comprehension scheme, and $\mathsf{ACA}\!\restriction$ and $\mathsf{ACA}^{\mathsf{pf}}\!\restriction$ their counterparts with induction involving only $\mathcal{L}$-formulas. From the proof of Proposition 1 we can thus extract:

Proposition 2 *(i)* $\mathsf{UTB}\!\restriction$ *interprets* $\mathsf{ACA}^{\mathsf{pf}}\!\restriction$
(ii) UTB *interprets* $\mathsf{ACA}^{\mathsf{pf}}$.
In both cases, the translation behaves like id *for arithmetical vocabulary.*

Proposition 2 tells us that the admission of second-order parameters in the comprehension schema corresponds, on the truth-theoretic side, to a substantial jump from a disquotational theory of truth to a compositional one. The substantiality of this jump can be measured in terms of proof-theoretic strength—and thus in terms of interpretability strength—as Lemma 6 shows. In other words, the presence of parameters in the comprehension schema enables one to go from a theory that is conservative over PA to a theory that proves $\mathsf{Con}(\mathsf{PA})$.

On the negative side, we have

Proposition 3 *(i)* $\mathsf{TB}\!\upharpoonright$, $\mathsf{UTB}\!\upharpoonright$, $\mathsf{CT}\!\upharpoonright$ *do not interpret* ACA_0*;*
(ii) CT^- *interprets* $\mathsf{ACA}_{\Pi^1_1}$

Proof (i) If one of $\mathsf{TB}\!\upharpoonright$, $\mathsf{UTB}\!\upharpoonright$, $\mathsf{CT}\!\upharpoonright$ interpreted ACA_0, by Lemma 7, PA would interpret ACA_0, quod non. (ii) follows from Cantini (1989, Proposition 3.6). □

Positive inductive definitions can be employed to define the set of arithmetical truths (cf. Moschovakis (1974)). The idea is to identify this set as the fixed point of a suitable operator on sets of natural numbers extracted by the inductive definition (cf. Halbach (2014, Sect. 8.7)). The theory PT captures the clauses of this inductive definition. PT and its subsystems are studied in Fischer (2009). The peculiarity of positive inductive definitions is that in their clauses the truth predicate can only appear in the scope of an even number of negation symbols.

Definition 3 (i) PT extends PAT with the axioms:

$$\mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \rightarrow \big(\mathsf{Tr}\, x \underset{\cdot}{\circ} y \leftrightarrow \mathsf{val}(x) \circ \mathsf{val}(y)\big) \tag{PT1}$$

$$\mathsf{CTerm}_{\mathcal{L}}(x) \wedge \mathsf{CTerm}_{\mathcal{L}}(y) \rightarrow \big(\mathsf{Tr}\, \underset{\cdot}{\neg}\, x \underset{\cdot}{\circ} y \leftrightarrow \neg\, (\mathsf{val}(x) \circ \mathsf{val}(y))\big) \tag{PT2}$$

$$\mathsf{Sent}_{\mathcal{L}}(x) \rightarrow (\mathsf{Tr} \underset{\cdot}{\neg} \underset{\cdot}{\neg} x \leftrightarrow \mathsf{Tr}\, x) \tag{PT3}$$

$$\mathsf{Sent}_{\mathcal{L}}(x \underset{\cdot}{\wedge} y) \rightarrow (\mathsf{Tr}\, x \underset{\cdot}{\wedge} y \leftrightarrow (\mathsf{Tr}\, x \wedge \mathsf{Tr}\, y)) \tag{PT4}$$

$$\mathsf{Sent}_{\mathcal{L}}(x \underset{\cdot}{\wedge} y) \leftrightarrow (\mathsf{Tr}\, \underset{\cdot}{\neg}(x \underset{\cdot}{\wedge} y) \leftrightarrow (\mathsf{Tr}\, \underset{\cdot}{\neg} x \vee \mathsf{Tr} \underset{\cdot}{\neg} y) \tag{PT5}$$

$$\mathsf{Sent}_{\mathcal{L}}(\underset{\cdot}{\forall} v x) \leftrightarrow (\mathsf{Tr}\, \underset{\cdot}{\forall} v x \leftrightarrow \forall y (\mathsf{Cterm}(y) \rightarrow \mathsf{Tr}(x(y/v)))) \tag{PT6}$$

$$\mathsf{Sent}_{\mathcal{L}}(\underset{\cdot}{\neg} \underset{\cdot}{\forall} v x) \leftrightarrow (\mathsf{Tr} \underset{\cdot}{\neg} \underset{\cdot}{\forall} v x \leftrightarrow \exists y (\mathsf{Cterm}(y) \wedge \mathsf{Tr} \underset{\cdot}{\neg} x(y/v))) \tag{PT7}$$

(ii) $\mathsf{PT}\!\upharpoonright$ restricts the induction scheme to $\mathcal{L}$–formulas
(iii) In PT^- the induction scheme of PAT is replaced by (IInd).

PT is identical to CT, as PT proves (CT2) by formal induction on the complexity of sentences. The proof of Proposition 1 carries over without modifications: PT thus interprets ACA in the manner described. $\mathsf{PT}\!\upharpoonright$ on the other hand, that is when induction is restricted to $\mathcal{L}$-sentences, is a *proper* subtheory of $\mathsf{CT}\!\upharpoonright$. To see this, one notices that the general theory of inductive definitions entails the existence of fixed points of positive inductive definitions: this fact can be used to define a subset $S \subseteq \mathcal{M}$ for any $\mathcal{M} \vDash \mathsf{PA}$ such that $(\mathcal{M}, S) \vDash \mathsf{PT}\!\upharpoonright$ (cf. Cantini 1989). By a theorem of Lachlan (1981), by contrast, not any model of PA can be expanded to a model of $\mathsf{CT}\!\upharpoonright$. This explains the properness of the subtheory relation. Therefore, by Lemma 7, PA interprets $\mathsf{PT}\!\upharpoonright$.

To get closer to our interests and consider subsystems of second-order arithmetic, we have:

Proposition 4 *(i)* PT *interprets* ACA*;*
(ii) PT^- *interprets* ACA_0.

Proof (i) follows from the identity of CT and PT and Lemma 1. (ii) is proved by Fischer (2009): the comprehension axioms require external induction on the complexity of the formula involved. Compare (14.5)–(14.7) in the proof of Proposition 1: in (14.6), the hidden induction on the complexity of formulas is replaced here by an induction on its *positive* complexity. Moreover, PT^- proves the translation of $\mathsf{tot}(x)$ where $x \in \mathsf{Fml}^1_{\mathcal{L}}$. □

14.3.2 Truth from Comprehension

Proposition 5 ACA *defines the truth predicate of* CT.

Proof The proof is contained in Takeuti (1987). Let $\mathsf{lcx}(\cdot)$ be as above, and $\mathcal{T}(X, x)$ mean 'X is a truth set for sentences of complexity $\leq x$', that is

$$\begin{aligned}\mathcal{T}(X, x) \;:\leftrightarrow\; & \forall u, v\, (\mathsf{Cterm}_{\mathcal{L}}(u) \wedge \mathsf{Cterm}_{\mathcal{L}}(v) \to (u \underset{\cdot}{\circ} v \in X \;\leftrightarrow\; \mathsf{val}(u) \circ \mathsf{val}(v))) \wedge \\ & \forall u\, (\mathsf{Sent}_{\mathcal{L}}(\underset{\cdot}{\neg} u) \wedge \mathsf{lcx}(\underset{\cdot}{\neg} u) \leq x \to (\underset{\cdot}{\neg} u \in X \;\leftrightarrow\; u \notin X)) \wedge \\ & \forall u, v\, (\mathsf{Sent}_{\mathcal{L}}(u \underset{\cdot}{\wedge} v) \wedge \mathsf{lcx}(u \underset{\cdot}{\wedge} v) \leq x \to (u \underset{\cdot}{\wedge} v \in X \leftrightarrow (u \in X \wedge v \in X))) \wedge \\ & \forall u, v\, (\mathsf{Sent}_{\mathcal{L}}(\underset{\cdot}{\forall} uv) \wedge \mathsf{lcx}(\underset{\cdot}{\forall} uv) \leq x \to (\underset{\cdot}{\forall} uv \in X \leftrightarrow \forall y (\mathsf{CTerm}_{\mathcal{L}}(y) \to \mathsf{sb}(y, v) \in X)))\end{aligned}$$

It is first noticed that $\mathsf{ACA}^{\mathsf{pf}}$ with the induction schema restricted to $\mathcal{L}_2$-formulas with no bound set variables proves, by induction on x, that truth definitions for sentences of restricted complexity are unique:

$$\mathcal{T}(X, x) \wedge \mathcal{T}(Y, x) \wedge \mathsf{Sent}_{\mathcal{L}}(y) \wedge \mathsf{lcx}(y) \leq x \to \;\; (y \in X \;\leftrightarrow\; y \in Y) \tag{14.8}$$

By arithmetical comprehension, one obtains, in $\mathsf{ACA}^{\mathsf{pf}}\!\restriction$,

$$\exists X\, \mathcal{T}(X, 0), \tag{14.9}$$

because sentences with logical complexity ≤ 0 are atomic statements. Moreover, in ACA_0, we have, still by arithmetical comprehension, that

$$\forall x\, (\exists X (\mathcal{T}(X, x)) \to \exists X (\mathcal{T}(X, x + 1))) \tag{14.10}$$

An instance of the induction schema of ACA (indeed $\mathsf{ACA}_{\Sigma^1_1}$ would suffice, as we only need an instance of Σ^1_1-induction), gives us

$$\forall x\, \exists X\, \mathcal{T}(X, x) \tag{14.11}$$

The required truth definition is given by the formula

$$\tau(x) :\leftrightarrow \exists X(\mathcal{T}(X, \mathsf{lcx}(x)) \wedge x \in X) \tag{14.12}$$

□

Corollary 1 *(i)* $\mathsf{ACA}^{\mathsf{pf}}\restriction$ *defines the truth predicate of* $\mathsf{UTB}\restriction$*;*
(ii) $\mathsf{ACA}^{\mathsf{pf}}$ *defines the truth predicate of* UTB*;*
(iii) $\mathsf{ACA}_{\Pi^1_1}$ *defines the truth predicate of* CT^-*;*
(iv) ACA *defines the truth predicate of* PT.

Proof (i) and (ii) are immediate by inspection of the proof of Proposition 5, by induction on the complexity of the formula involved in the uniform diquotation. (iii), in addition, requires Π^1_1 induction to prove CT2. (iv) is immediate by Proposition 5 and the fact that PT is a subtheory of CT. □

14.3.3 Extensions and Hierarchies

The processes of extending PA to CT or ACA can be iterated. To this end, one first assumes a suitable notation for ordinals below Γ_0 (cf. Pohlers 2009, Chap. 2): in other words, any ordinal $\alpha < \Gamma_0$ can be coded by a natural number and thus represented in PA by a unique term $\overline{\alpha}$.

To define the theories $\mathsf{RT}_{<\theta}$ for $\theta \leq \Gamma_0$—RT standing for 'ramified truth'—one considers languages $\mathcal{L}_{<\theta} := \mathcal{L}_{\mathsf{PA}} \cup \{\mathsf{T}_0, \ldots, \mathsf{T}_\eta\}$ for $\eta < \theta$, where $\mathsf{T}_\alpha := (\mathsf{T}, \alpha)$. The axioms of $\mathsf{RT}_{<\theta}$ thus stipulate how the truth predicates T_η work for sentences of $\mathcal{L}_{<\eta}$, for $\eta < \theta$ (cf. Halbach 2014, p. 113).

The definition of iterations of ACA to $\theta \leq \Gamma_0$, called $\mathsf{RA}_{<\theta}$ for 'ramified analysis up to θ', enjoys different variations. Since we are not so much interested in iterations of predicative comprehension, we refer to Feferman (1964) for the definition of $\mathsf{RA}_{<\theta}$.

The fundamental fact that links ramified truth and analysis is the following, whose proof is sketched in Feferman (1991):

Proposition 6 (Feferman) *For* $\alpha \leq \Gamma_0$*,* $\mathsf{RT}_{<\alpha}$ *and* $\mathsf{RA}_{<\alpha}$ *are mutually interpretable.*

In the light of Proposition 6, one may for instance obtain a reduction of the classical axiomatization of Kripke's theory of truth KF (cf. Feferman 1991, Halbach 2014) to $\mathsf{RT}_{<\epsilon_0}$: in particular, KF defines all truth predicates of $\mathsf{RT}_{<\epsilon_0}$. Also, from Cantini (1989, Sect. 9) one can extract an interpretation of KF in $\mathsf{RT}_{<\epsilon_0}$. Given the mutual truth definability between the remarkable disquotational system PUTB of positive uniform (type free) disquotation and KF (cf. Halbach 2009), Proposition 6 closely relates PUTB and iterations of predicative comprehension. It also shows the close relationships between KF, PUTB, and $\widehat{\mathsf{ID}}_1$.[10]

[10] $\widehat{\mathsf{ID}}_1$ is the theory postulating the existence of fixed-points for arbitrary positive arithmetical operators. For a detailed definition of the full ID_1, cf. Pohlers (2009, Chap. 9, Definition 9.1).

Many other results and reductions between type-free truth and second order-arithmetic have been investigated in recent years. As we anticipated in the introduction, however, the strategy put forward in the next section does not enjoy a plausible extension to the type-free case yet: we lack natural ways of extending the strategy proposed in the next section allow self-applications of the operations on theories to be defined. Therefore we decided to omit a systematic survey of the relationships between type-free systems and subsystems of second-order arithmetic.

14.4 Truth and Predicative Comprehension as Functors

As anticipated in the introduction, we will now consider the reductions surveyed in Sect. 14.3 from a different angle. This approach finds its roots in a philosophical attempt to distinguish patterns of reasoning belonging to the syntactico/truth-theoretic component of a theory of truth on one side and its mathematical/object-theoretic component on the other. This separation has been used, also by the author, to analyze the connections between axiomatic truth and truth-theoretic deflationism (cf. Halbach 2014, Heck 2015, Nicolai 2015b). The fundamental insight offered by this approach is thus that, unlike what happens in the standard construction, the assumption of a full, Tarskian theory of truth does not immediately lead to new object-theoretic, or mathematical insights, although it may do so via a suitably defined interpretation.

In Sect. 14.4.1 we define and introduce the basic properties of $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$ that operate on theories U by applying Tarskian truth and positive, typed truth to them. Although inductive reasoning involving truth-theoretic and syntactic notions is not directly available in $\mathsf{T}[\cdot]$ (and $\mathsf{Tp}[\cdot]$), it can be mimicked in it by exploiting the properties of definable cuts introduced in Sect. 14.2: this will be the core of Sect. 14.4.2. The arithmetized model corresponding to the formalization—in weak arithmetical context—of the construction of the term model used in Henkin's proof of the completeness theorem, comes with a truth predicate. This truth predicate is used in Sect. 14.4.3 to interpret the truth predicates of $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$. In Sect. 14.4.4 we turn to predicative comprehension and investigate the functors $\mathsf{PC}[\cdot]$ and $\mathsf{PCS}[\cdot]$; in Sect. 14.4.5 we finally relate them to $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$.

In the following sections the object theory U will be assumed to be formulated in a relational language. On occasion we will require sequentiality or the capability of interpreting S^1_2.

14.4.1 Typed-Truth as an Operation on Theories

The functor $\mathsf{T}[\cdot]$ applies to an arbitrary RE theory U and yields the three-sorted theory $\mathsf{T}[U]$ in a language $\mathcal{L}_T$ with sorts in $\{\mathfrak{s}, \mathfrak{o}, \mathfrak{sq}\}$ (for 'syntax', 'object-theory', 'sequences/variable assignments'), constants proper of the language $\mathcal{L}_{\#}$ of S^1_2,[11] the function symbol $\cdot(\cdot)$ of type $(\mathfrak{sq}, \mathfrak{s}) \to \mathfrak{o}$ and the predicate symbol Sat of sort $(\mathfrak{s}, \mathfrak{sq})$. The former will give rise to expressions of the form $v_i^{\mathfrak{s}q}(v_i^{\mathfrak{s}}) = v_i^{\mathfrak{o}}$, stating that the $v_i^{\mathfrak{s}}$th element of a variable assignment $v_i^{\mathfrak{s}q}$ is the U-object $v_i^{\mathfrak{o}}$, whereas the latter will be characterized as a satisfaction predicate. Greek letters $\varphi, \psi, \ldots$ are taken to range over formulas of $\mathcal{L}_U$.

The axioms of $\mathsf{T}[U]$, besides the axioms of U, are

$$\text{axioms of } \mathsf{S}^1_2 \tag{S^1_2}$$

$$\exists v_j^{\mathfrak{sq}}\,(\forall v_k^{\mathfrak{s}}(v_k^{\mathfrak{s}} \neq v_i^{\mathfrak{s}} \to v_i^{\mathfrak{sq}}(v_k^{\mathfrak{s}}) = v_j^{\mathfrak{sq}}(v_k^{\mathfrak{s}})) \wedge v_j^{\mathfrak{sq}}(v_i^{\mathfrak{s}}) = v_i^{\mathfrak{o}}) \tag{sq}$$

$$\begin{array}{l}\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner R(v_1^{\mathfrak{o}}, \ldots, v_n^{\mathfrak{o}})\urcorner) \leftrightarrow R(v_i^{\mathfrak{sq}}(\ulcorner v_1^{\mathfrak{o}}\urcorner), \ldots, v_i^{\mathfrak{sq}}(\ulcorner v_n^{\mathfrak{o}}\urcorner)) \\ \text{for every relation symbol } R \text{ in } \mathcal{L}_U\end{array} \tag{tat}$$

$$\mathsf{Sat}(v_i^{\mathfrak{sq}}, \underset{\cdot}{\neg}\ulcorner\varphi\urcorner) \leftrightarrow \neg\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi\urcorner) \tag{t$\neg$}$$

$$\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi \vee \psi\urcorner) \leftrightarrow \mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi\urcorner) \vee \mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\psi\urcorner) \tag{t$\vee$}$$

$$\mathsf{Sat}(v_i^{\mathfrak{s}q}, \ulcorner\exists v_i^{\mathfrak{o}}\varphi\urcorner) \leftrightarrow (\exists v_j^{\mathfrak{sq}} \overset{\ulcorner v_i^{\mathfrak{o}}\urcorner}{\sim} v_i^{\mathfrak{sq}})(\mathsf{Sat}(v_j^{\mathfrak{sq}}, \ulcorner\varphi\urcorner)) \tag{tq}$$

where

$$v_j^{\mathfrak{sq}} \overset{v_i^{\mathfrak{s}}}{\sim} v_i^{\mathfrak{sq}} \;\leftrightarrow\; \forall v_j^{\mathfrak{s}}(v_j^{\mathfrak{s}} \neq v_i^{\mathfrak{s}} \to v_j^{\mathfrak{sq}}(v_j^{\mathfrak{s}}) = v_i^{\mathfrak{sq}}(v_j^{\mathfrak{s}}))$$

(sq) states minimal conditions on the existence and on the behaviour of sequences. (tq) simply formalizes Tarski's clause on quantified formulas: $\exists v_i\varphi$ is satisfied by a variable assignment if and only if there is a second variable assignment, differing from the former only in what it assigns to v_i (this is captured by the abbreviation $v_j^{\mathfrak{sq}} \overset{v_i^{\mathfrak{s}}}{\sim} v_i^{\mathfrak{sq}}$), satisfying φ.

Intuitively, by applying the operator $\mathsf{T}[\cdot]$ to U, one expands a model of U with a universe of syntactic objects—here numbers in the sense of S^1_2—and a universe of 'mixed' objects, sequences of $\mathcal{L}_U$-variable assignments living in a disjoint domain.

We define an alternative functor: $\mathsf{Tp}[U]$ adds *positive* compositional axioms similar to the axioms of $\mathsf{PT}\!\restriction$ in Definition 3 but in a many-sorted way compatible with the idea behind the construction of $\mathsf{T}[U]$. Again $\mathsf{Tp}[U]$ is formulated in $\mathcal{L}_T$; its axioms, besides the axioms of U, S^1_2, (sq), (tat), (t$\vee$), are

$$\begin{array}{l}\mathsf{Sat}(v_i^{\mathfrak{sq}}, \underset{\cdot}{\neg}\ulcorner R(v_1^{\mathfrak{o}}, \ldots, v_n^{\mathfrak{o}})\urcorner) \leftrightarrow \neg(R(v_i^{\mathfrak{sq}}(\ulcorner v_1^{\mathfrak{o}}\urcorner), \ldots, v_i^{\mathfrak{sq}}(\ulcorner v_n^{\mathfrak{o}}\urcorner))) \\ \text{for every relation symbol } R \text{ in } \mathcal{L}_U\end{array} \tag{fat}$$

[11] For simplicity, we may require them to be unofficial abbreviations of their relational counterparts.

$$\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg}\dot{\neg}\ulcorner\varphi\urcorner) \leftrightarrow \mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi\urcorner) \tag{tdn}$$

$$\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg}\ulcorner\varphi \vee \psi\urcorner) \leftrightarrow \big(\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg}\ulcorner\varphi\urcorner) \wedge \mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg}\ulcorner\psi\urcorner)\big) \tag{f$\vee$}$$

$$\mathsf{Sat}(v_i^{\mathfrak{sq}}, \dot{\neg}\ulcorner\exists v_i^{\mathfrak{o}}\varphi\urcorner) \leftrightarrow (\forall v_j^{\mathfrak{sq}} \overset{\ulcorner v_i^{\mathfrak{o}}\urcorner}{\sim} v_i^{\mathfrak{sq}})(\mathsf{Sat}(v_j^{\mathfrak{sq}}, \dot{\neg}\ulcorner\varphi\urcorner)) \tag{fq}$$

By external induction on the positive complexity of $\varphi(\vec{v_i^{\mathfrak{o}}})$, we have:

Lemma 8 $\mathsf{Tp}[U]$ *(and thus* $\mathsf{T}[U]$*) proves, for all* $\varphi(\vec{v_i^{\mathfrak{o}}})$ *in* $\mathcal{L}_U$*,*
$\mathsf{Sat}(v_i^{\mathfrak{sq}}, \ulcorner\varphi(v_{i_1}^{\mathfrak{o}}, \ldots, v_{i_n}^{\mathfrak{o}})\urcorner) \leftrightarrow \varphi(v_i^{\mathfrak{sq}}(\ulcorner v_{i_1}^{\mathfrak{o}}\urcorner), \ldots, v_i^{\mathfrak{sq}}(\ulcorner v_{i_n}^{\mathfrak{o}}\urcorner))$

We will often refer to variables of sort $\mathfrak{o}, \mathfrak{s}, \mathfrak{sq}$ as $u, v, w, x, y, z \ldots,\ i, j, k, l, m, n \ldots$, and $a, b, c, \ldots$ respectively.

14.4.2 Consistency from Truth

A remarkable fact concerning the method of shortening of cuts, introduced in Sect. 14.2, is that it enables one to prove the consistency of the object theory U relativized to a definable cut, once we are able to prove, in the theory of truth, that all axioms of U are true.

Therefore we define

$$\mathsf{AxT}_U :\leftrightarrow \forall a \forall k\, (\mathsf{AxL}_U(k) \vee \mathsf{Ax}_U(k) \rightarrow \mathsf{Sat}(a, k))$$

where $\mathsf{AxL}_U(k)$ and $\mathsf{Ax}_U(k)$ are Δ_1^b-representations of the logical and the nonlogical axioms of U in S_2^1. AxT_U thus states that all logical and nonlogical axioms of U are true, i.e. satisfied by all sequences. This definition of AxT_U enables us to simplify the proof of a similar result given in Nicolai (2015a); to this extent, we recall that U is taken to be formulated in a Hilbert-style calculus in which Modus Ponens is the only rule of inference. It should be noticed, however, that the result is not dependent on the choice of a specific logical calculus: this only simplifies the description. Thus we omit details on the specific formulation of U in the statements of the results.

To simplify the notation, from now on we set:

$$\mathsf{T}^+[U] := \mathsf{T}[U] + \mathsf{AxT}_U$$
$$\mathsf{Tp}^+[U] := \mathsf{Tp}[U] + \mathsf{AxT}_U$$

Lemma 9 $\mathsf{T}[U]^+$ *proves the consistency of U on a cut.*

Proof We recall that $\mathsf{Prf}_U(k)$ is taken to be a Δ_1^b-formula in S_2^1 expressing that k is the code of a U-proof, and $\mathsf{lst}(k)$ a Σ_1^b function yielding, when applied to a sequence k, the last element of k.

We break down the proof in several steps.

We first consider the following formula with only n free.

$$\mathcal{K}(n) :\leftrightarrow \ (\forall m \leq n) \ \big(\mathsf{Prf}_U(m) \to \forall a \ \mathsf{Sat}(a, \mathsf{lst}(m))\big) \tag{14.13}$$

It expresses that any proof smaller than n has a true conclusion. We show that it is inductive. The claim clearly holds for $n = 0$, by assumption. Assuming it holds for arbitrary proofs $k, l \leq n$ and sequence a, we want to prove it for proofs $m \leq n + 1$. In the interesting case, we have

$$\mathsf{lst}(m) = \mathsf{mp}(\mathsf{lst}(k), \mathsf{lst}(l)),$$

where $\mathsf{mp}(m, n)$ is a Σ_1^b-function yielding the result of applying Modus Ponens to m and n. In particular, here we require $\mathsf{lst}(k)$ of the form $\mathsf{lst}(l)\dot{\to}\mathsf{lst}(m)$. The claim is thus obtained by applying (t¬) and (t∨) to $\mathsf{Sat}(a, \mathsf{lst}(k))$ and combining it with $\mathsf{Sat}(a, \mathsf{lst}(l))$.

We shorten $\mathcal{K}$ to a cut $\mathcal{J}$ such that, for formulas j, we have

$$\exists m \ (\mathcal{J}(m) \wedge \mathsf{Proof}_U(m, j)) \to \forall a \ \mathsf{Sat}(a, j)$$

We can safely assume that S_2^1 is available in $\mathcal{J}$.

Assuming the monotonicity of the coding—that is codes of elements of sequences are smaller of the code of the whole sequence—if $m \in \mathcal{J}$, also j will be. Finally we reason in the standard way: let $\perp$ code, in the S_2^1-numbers, an absurdity in the sense of U. If $\mathsf{Pr}_U^{\mathcal{J}}(\perp)$, then $\mathsf{Sat}(a, \perp)$. By Lemma 8, we obtain $\neg\mathsf{Pr}_U^{\mathcal{J}}(\perp)$, that is $\neg \exists m \ (\mathcal{J}(m) \wedge \mathsf{Proof}_U(m, \perp))$, that is $\mathsf{Con}^{\mathcal{J}}(U)$. □

By Lemma 8, the parametrized Tarski-biconditionals are provable in $\mathsf{Tp}[U]$. To prove that all logical axioms of A are true on a cut, we need extra-care. In a word, we have to make sure that syntax and truth interact well on this cut. The problem is essentially that, in order to prove the truth of a logical axiom schema, say $\forall v_0 \varphi \to \varphi(y/v_0)$, one has to employ an instance of the very same principle, e.g.

$$\forall m \mathsf{Sat}(a, k(m/\ulcorner v_0 \urcorner)) \to \mathsf{Sat}(a, k(t/\ulcorner v_0 \urcorner)) \tag{14.14}$$

to be able to apply (t¬),(t∨), (tq) and obtain

$$\mathsf{Sat}(a, \ulcorner \forall v_0 \varphi \to \varphi(y/v_0) \urcorner) \tag{14.15}$$

The step from (14.14) to (14.15) requires, for instance, that variable assignments and formal substitution behave as required under the scope of Sat. Since these lemmata are usually proved by induction, we need to resort to a definable cut in which this behaviour is preserved. For details, we refer to Heck (2015) and Nicolai (2015a). Once this is done, we have (recall that $A, B, C \ldots$ range over finitely axiomatized theories):

Corollary 2 $\mathsf{T}[A]$ *proves the consistency of A on a cut.*

Surprisingly enough, a similar results can be transferred to $\mathsf{Tp}^+[U]$. We have already seen that CT and PT, unlike their variants featuring arithmetical induction only, are extensionally identical. Induction extended to the truth predicate is in fact needed to prove the full axiom for negation. The technology of shortening of cuts enables us to mimic the role played by the extended induction schema of PT in deriving (CT2).

Lemma 10 *There is a* $\mathsf{Tp}[U]$*-definable cut in which (t¬) holds.*

Proof We reason in $\mathsf{Tp}[U]$. Let $\mathsf{lc}(x)$ a Σ^b_1-function that, applied to a $\mathcal{L}_U$-formula φ, yields the number of its logical symbols. By Lemma 1 the claim is obtained once we show that the following is inductive:

$$\mathcal{K}_0(n) := \forall k \forall a(\mathsf{Fml}_{\mathcal{L}_U}(k) \wedge \mathsf{lc}(k) \leq n \rightarrow (\mathsf{Sat}(a, \dot{\neg} k) \leftrightarrow \neg \mathsf{Sat}(a, k))$$

If $n = 0$, we use (tat) and (fat). Assuming the claim is true for formulas k with $\mathsf{lc}(k) \leq n$, we obtain the claim by the truth/falsity clauses for connectives and quantifiers and the 'inductive' assumption. For instance, if $k = \dot{\neg} k_0$, we have:

$$\begin{aligned} \neg \mathsf{Sat}(a, \dot{\neg} k_0) &\leftrightarrow \neg\neg \mathsf{Sat}(a, k_0) && \text{by assumption} \\ &\leftrightarrow \mathsf{Sat}(a, k_0) && \\ &\leftrightarrow \mathsf{Sat}(a, \dot{\neg}\dot{\neg} k_0) && \text{(tdn)} \end{aligned}$$

□

Corollary 3 $\mathsf{Tp}[U]$ *interprets* $\mathsf{T}[U]$ *on a definable cut.*

More generally, Lemma 10 gives a strategy to interpret $\mathsf{CT}\restriction$ into $\mathsf{PT}\restriction$, the standard systems of Tarskian and positive truth considered in Sect. 14.3.

Once we have (t¬) available on suitable initial segment $\mathcal{K}_0$, we can run the proof of Lemma 9 paying attention to work in $\mathsf{Tp}^+[U]$-definable cuts that shorten $\mathcal{K}_0$ and in which the axioms of S^1_2 are satisfied. We thus have:

Proposition 7 $\mathsf{Tp}[U]^+$ *proves the consistency of* U *on a cut.*

Proof Instead of $\mathcal{K}(n)$ in (14.13), we can start with the formula

$$\mathcal{K}_1(n) :\leftrightarrow (\forall m \leq n)(\mathsf{Prf}_U(m) \wedge \mathcal{K}_0(\mathsf{lst}(m)) \rightarrow \forall a\, \mathsf{Sat}(a, \mathsf{lst}(m))) \tag{14.16}$$

By the same argument used in the case of $\mathcal{K}(n)$ in the proof of Lemma 9, it can be shown that $\mathcal{K}_1(n)$ is inductive (only here the properties of $\mathcal{K}_0$ are used in the inductive step). The proof then proceeds unchanged. □

By Lemma 8, and taking care of truth of all logical axioms of A on a suitable definable cut as suggested above, we have

Corollary 4 $\mathsf{Tp}[A]$ *proves the consistency of* A *on a cut.*

Finally, we state the desired property of the result of applying $\mathsf{T}^+[\cdot]$ and $\mathsf{Tp}^+[\cdot]$ to arbitrary base theories:

Corollary 5 *(i)* $\mathsf{T}^+[U]$ *and* $\mathsf{Tp}^+[U]$ *interpret* $\mathsf{S}^1_2+\mathsf{Con}(U)$*, and thus* $\mathsf{Q}+\mathsf{Con}(U)$*; (ii)* $\mathsf{T}[A]$ *and* $\mathsf{Tp}[A]$ *interpret* $\mathsf{S}^1_2+\mathsf{Con}(A)$*, and thus* $\mathsf{Q}+\mathsf{Con}(A)$.

Proof Ad (i), we simply take the cut $\mathcal{J}$ and the cut resulting from $\mathcal{K}_1$ from Proposition 7 as domains of our interpretations. We notice that, by Lemma 2, S^1_2 is holds in them. (ii) is a weaker claim than (i). □

14.4.3 Truth from Consistency

We now turn to the question of how to recover the semantic machinery of $\mathsf{T}[\cdot]$ and $\mathsf{Tp}[\cdot]$ via the assumption of the consistency of the base theory. Feferman offered in Feferman (1960) a full formalization of Henkin's term model construction for a first-order, recursive, set of sentences S in extensions of PA plus a Π_1 assertion of the consistency of S. Following Visser (1991, 2009), we refer to the resulting arithmetized model as the 'Henkin–Feferman construction'. The technology of definable cuts enables us to formalize this term model in very weak subsystems of arithmetic complemented with the consistency statement for S.

Proposition 8 $\mathsf{S}^1_2+\mathsf{Con}(U)$ *interprets* U.

Proof Sketch A full proof can be found in Visser (1991). As usual, we start with $\mathcal{L}_U$ and extend it to $\mathcal{L}^h_U$ by adding a constant $c_{\exists v_i\varphi}$ for every $\exists v_i\varphi$ in the extended language: notice that we have enough induction in $\mathsf{S}^1_2+\mathsf{Con}(U)$ to formalize this inductive argument. The Henkin theory, from which the arithmetized model for U is read off, is also defined in stages:

$$U^h_0 := U$$

$$U^h_{n+1} := \begin{cases} U^h_n \cup \{\varphi\}, & \text{if } n = \ulcorner\varphi\urcorner \text{ (with } \varphi\in\mathcal{L}^h_U\text{) and } \mathsf{Con}(U^h_n\cup\{\varphi\}),\\ U^h_n\cup\{\varphi\}\cup\{\psi(c_\varphi)\}, & \text{if } \varphi \text{ is } \exists v\psi \text{ and } \mathsf{Con}(U^h_n\cup\{\varphi\})\\ U^h_n & \text{otherwise}\end{cases}$$

Unlike the definition of $\mathcal{L}^h_U$, Σ^b_1-induction is not sufficient to prove the existence of the union of all U^h_n's. However, it turns out that the formula defining the U^h_n's is inductive, so it can be shortened to a cut $\mathcal{I}$. The required, complete theory in $\mathcal{L}^h_U$, for sentences in $\mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$ can thus be taken to be $\mathcal{F}=\bigcup_{n\in\mathcal{I}}U^h_n$.

From $\mathcal{F}$, we define the required interpretation $\mathfrak{F}$, whose domain $\delta_{\mathfrak{F}}$ is the set of codes of Henkin constants of $\mathcal{L}^h_U$ in $\mathcal{I}$. To the theory $\mathcal{F}$ there corresponds a truth predicate S. We set

$$R^{\mathfrak{F}}(v^{\mathrm{o}}) :\leftrightarrow \delta_{\mathfrak{F}}(x)\wedge R(x)\in\mathsf{S} \qquad \text{for any } R\in\mathcal{L}_U \tag{14.17}$$

We notice that in (14.17) and in what follows, we write $\varphi(x)\in S$ to express that the result of formally substituting v_1 for x in φ falls into the extension of S. Crucially, $\mathsf{S}^1_2+\mathsf{Con}(U)$ proves:

(a) for $\varphi \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$, $(\mathsf{S}\ulcorner\neg\varphi\urcorner) \leftrightarrow \neg(\mathsf{S}\ulcorner\varphi\urcorner)$;

(b) for $\varphi, \psi \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$, $\mathsf{S}(\ulcorner\varphi \vee \psi\urcorner) \leftrightarrow (\mathsf{S}(\ulcorner\varphi\urcorner) \vee \mathsf{S}(\ulcorner\psi\urcorner))$;

(c) for $\exists v\varphi(v) \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}^h_U}$ and $v^{\mathfrak{o}} \in \mathsf{Var}^{\mathcal{I}}$, $\mathsf{S}(\ulcorner\exists v^{\mathfrak{o}}\varphi\urcorner) \leftrightarrow \exists x\,(\delta_{\mathfrak{F}}(x) \wedge (\varphi(x) \in \mathsf{S}))$;

(d) for all $\varphi \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}_U}$, $(\mathsf{Pr}_U(\ulcorner\varphi\urcorner) \to \mathsf{S}(\ulcorner\varphi\urcorner))$

Notice, in (d), that reflection holds for sentences of the *non extended language* $\mathcal{L}_U$ in the cut $\mathcal{I}$. □

By the interpretability of $\mathsf{S}^1_2 + \mathsf{Con}(U)$ in $\mathsf{Q} + \mathsf{Con}(U)$ on a definable cut, we have:

Corollary 6 $\mathsf{Q} + \mathsf{Con}(U)$ *interprets* U.

Now we show that the truth predicate $\mathsf{S}(\cdot)$ defined in the proof of Proposition 8 enables us to interpret the satisfaction predicate of $\mathsf{T}^+[U]$, and thus of $\mathsf{Tp}^+[U]$. Therefore we obtain Tarskian and positive, typed truth for the base theory from the assertion of the consistency of U.

Proposition 9 $\mathsf{S}^1_2 + \mathsf{Con}(U)$ *interprets* $\mathsf{T}^+[U]$.

Proof We define the translation $\mathfrak{H}$. It is assume a suitable renaming of bound variables omitted for readability; S is the truth predicate defined in the proof of Proposition 8.

$$(R)^{\mathfrak{H}}(u) :\leftrightarrow R(u) \in \mathsf{S} \quad \text{for predicates } \mathsf{R} \text{ of sort } \mathfrak{o} \tag{14.18}$$

$$(P)^{\mathfrak{H}}(u) :\leftrightarrow P(u) \quad \text{for predicates } P \text{ of sort } \mathfrak{s}; \tag{14.19}$$

$$(\mathsf{Sat})^{\mathfrak{H}}(u, v) :\leftrightarrow \mathsf{S}(\mathsf{sb}(v, u)); \quad \text{where } \mathsf{sb} \text{ substitutes elements of } u \text{ in } v \tag{14.20}$$

$$((\cdot)\cdot)^{\mathfrak{H}}(u, v, w) :\leftrightarrow ((u)_v = w \wedge v < \mathsf{lh}(u)) \vee (v \geq \mathsf{lh}(u) \wedge w = 0) \tag{14.21}$$

$$(\exists x\, A)^{\mathfrak{H}} :\leftrightarrow \exists x\ (\delta_{\mathfrak{F}}(x) \wedge A) \quad \text{where } \delta_{\mathfrak{F}} \text{ is as in Proposition 8;} \tag{14.22}$$

$$(\exists k\, A)^{\mathfrak{H}} :\leftrightarrow \exists k\ (\mathcal{I}(k) \wedge A) \quad \text{with } \mathcal{I} \text{ again as in Proposition 8} \tag{14.23}$$

$$(\exists a\, A)^{\mathfrak{H}} :\leftrightarrow \exists a, a \text{ is a (finite) sequence and } (\forall y \in a)(\delta_{\mathfrak{H}}(y)) \tag{14.24}$$

In (14.21), $(x)_y$ is an efficient version of the β function that outputs the yth element of the finite sequence x (cf. Buss 1986, 1998). In (14.22)–(14.24) we have kept the suggestive notation concerning variables even though, strictly speaking, we are not considering the language $\mathcal{L}_T$ anymore.

A full verification that $\mathfrak{H}$ is indeed a relative interpretation can be found in Nicolai (2015a). Crucially, for AxT_U, one notices that, by Proposition 8(d),

$$\forall u\ (\mathsf{Sent}_{\mathcal{L}_U}(u) \wedge \mathcal{I}(u) \wedge \mathsf{Ax}_U(u) \to \mathsf{S}(u)) \tag{14.25}$$

Therefore, it suffices to notice that in $\mathsf{S}^1_2 + \mathsf{Con}(U)$

$$\text{for all } u \in \mathsf{Sent}^{\mathcal{I}}_{\mathcal{L}_U}, \mathsf{Ax}^{\mathfrak{H}}_U(u) \to \mathsf{Ax}_U(u). \tag{14.26}$$

□

Corollary 7 *(i)* $\mathsf{S}^1_2 + \mathsf{Con}(U)$ *interprets* $\mathsf{Tp}^+[U]$*;*
(ii) $\mathsf{Q} + \mathsf{Con}(U)$ *interprets* $\mathsf{T}^+[U]$*, and thus* $\mathsf{Tp}^+[U]$.

Proof (i) is immediate from the previous Proposition. (ii) follows from Lemma 3. □

Corollary 7 clearly holds for $\mathsf{T}[A]$ and $\mathsf{Tp}[A]$.

14.4.4 The Functors PC[·] *and* PCS[·]

The facts reported in this subsection—except the connections between these facts and the results above—are essentially due to Visser (2009). We show that, given an arbitrary theory U,

The functor $\mathsf{PC}[\cdot]$ adds predicative comprehension to an arbitrary U. In order to be completely general, as our overall program requires, we need to be able to act on arbitrary U. To mention a familiar example, we should be able to generalize the transformation of PA into ACA_0 to arbitrary theories. This can be done in the following way.

We first render the language two-sorted: objects over which quantifiers of $\mathcal{L}_U$ range are taken to be of sort $\mathfrak{o}$. We add to the language variables X of sort $\mathfrak{c}$, standing for sets (or concepts, or unary predicates) of objects of sort $\mathfrak{o}$ and a binary predicate A of type $(\mathfrak{c}, \mathfrak{o})$ such that $Xx :\leftrightarrow \mathsf{A}(X, x)$. To obtain $\mathsf{PC}[U]$, the axioms of U formulated in the new language are extended with the scheme of predicative comprehension that, in a pedantic version, reads

$$\exists u^{\mathfrak{c}}\, \forall v^{\mathfrak{o}}\, (\mathsf{A}(u^{\mathfrak{c}}, v^{\mathfrak{o}}) \leftrightarrow \varphi(v^{\mathfrak{o}}, \vec{w}^{\mathfrak{o}}, \vec{y}^{\mathfrak{c}})) \tag{ca}$$

If U is sequential, then $\mathsf{PC}[U]$ can be finitely axiomatized (Visser 2009, Theorem 3.4).

In order to describe the next results in full generality, a slight modification of $\mathsf{PC}[\cdot]$ has to be considered. The required parallelism between consistency and predicative comprehension will break down if we considered arbitrary representations of the axiom set of U. A powerful theorem of Vaught states that any RE, sequential theory U can be axiomatized by a scheme. This scheme can be taken to be of the form $\Psi(\vec{B})$, where the B's are *schematic variables* for formulas of $\mathcal{L}_U$. Vaught's result states that U and the theory U_{V} resulting from the process just described are extensionally identical (cf. Vaught 1967).

To define the theory $\mathsf{PCS}[U]$, we formulate U_{V} in a language with sorts in $\{\mathfrak{o}, \mathfrak{c}\}$ and translate $\Psi(\vec{B})$ into $\Psi(\vec{X})$: we call this 'translation' of U_{V} as given by $\Psi(\vec{X})$ the *v-class form* of U and denote it with U^{c}. To obtain $\mathsf{PCS}[U]$, one then adds to U^{c} both (ca) and the universal closure of $\Psi(\vec{X})$. A modification of the strategy required in Lemma 1 shows that we can find a definable $\mathsf{PCS}[U]$-cut on which there are no proofs of U-contradictions. In other words:

Lemma 11 *Let* U_{V} *be sequential. Then* $\mathsf{PCS}[U]$ *proves the consistency of* U_{V} *on a cut.*

Proof Sketch We recall the main steps of Theorem 7.1 in Visser (2009), and reason in PCS[U]. Let $\mathsf{lc}(\cdot)$ be as above. By the sequentiality of U_{V}, syntax is formalized in a standard way (e.g. by the natural numbers). We consider a reformulation modification of the definition of $\mathcal{T}(X, x)$ in Proposition 5, which we call $\mathcal{T}^*(X, x)$.

Now in Lemma 5 we were able to use (ca) and the induction schema of ACA that the uniqueness condition for truth sets. In $\mathsf{PCS}[U]$ we have no such induction. We have to use again the shortening techniques. In particular, we can show that the (pseudo-) class

$$\mathcal{M} := \{x \mid \exists! X\ \mathcal{T}^*(X, x)\}, \tag{14.27}$$

that is the class of all numbers which bound the complexity of formulas occurring in our truth sets, is indeed inductive. For any x, this unique X is denoted with X^x. We shorten $\mathcal{M}$ to a cut $\mathcal{N}$.

A satisfaction predicate is now extracted from $\mathcal{N}$. $\mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$ is the class of formulas of $\mathcal{L}_U$, lying in $\mathcal{N}$, with one free variable. It is

$$\mathsf{St}(s, \ulcorner\varphi\urcorner) \;:\leftrightarrow\; (\exists x \in \mathcal{N})((s, \ulcorner\varphi\urcorner) \in X^x) \tag{14.28}$$

St enjoys the nice property that, for any φ in $\mathsf{Fml}^{1,\mathcal{J}}_{\mathcal{L}_U}$ and assignment s, one can find—again by (ca)—a Y_φ such that, for all y and v_i,

$$Y_\varphi y \text{ if and only if } \mathsf{St}(s(y/v_i), \ulcorner\varphi\urcorner) \tag{14.29}$$

In fact this Y_φ can simply be taken to be the set of y that satisfy φ in the sense of $X^{\mathsf{lc}(\varphi)}$. Since $\varphi \in \mathsf{Fml}^{1,\mathcal{J}}_{\mathcal{L}_U}$ we also know that Y_φ exists.

By employing an argument similar to the one employed in Lemma 9, we consider the set of U_{V}-proofs $\mathcal{P}$ whose members $\pi \in \mathcal{P}$ are such that $\mathsf{lh}(\pi) \leq x$, $\forall i \leq \mathsf{lh}(\pi)$, $(\pi)_i \in \mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$ and such that $\forall s\ \mathsf{St}(s, \mathsf{lst}(\pi))$. The set $\mathcal{A}$ of such x is inductive: crucially, for the axiom $\Psi(\vec{\psi})$ of U_{V} with $\vec{\psi} \in \mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$, we require

$$\mathsf{St}(s(y/v_i), \ulcorner\Psi(\vec{\psi})\urcorner) \tag{14.30}$$

By definition of PCS[U], we have $\forall\vec{Y}\ \Psi(\vec{Y})$ and thus $\Psi(\vec{\psi})$ for an arbitrary $\psi \in \mathsf{Fml}^{1,\mathcal{N}}_{\mathcal{L}_U}$. By (14.29), we associate to each formula in $\vec{\psi}$ the appropriate set Y_{ψ_i}. Thus $\Psi(\vec{Y}_\psi)$. By the properties of St, we obtain (14.30). A similar argument holds for the logical axiom schemata. The inductive step is unproblematic and it is similar to the inductive step in the proof of Lemma 9.

We thus shorten $\mathcal{A}$ to a cut $\mathcal{A}_0 \subset \mathcal{A}$. It follows that, as in Lemma 9, no proof of contradiction can be reached in $\mathcal{A}_0$: i.e. we have $\mathsf{Con}^{\mathcal{A}_0}(U_{\mathsf{V}})$. □

Corollary 8 *Let* U_{V} *be sequential. Then* $\mathsf{PCS}[U]$ *interprets* $\mathsf{S}^1_2 + \mathsf{Con}(U_{\mathsf{V}})$.

We now show, with the help of Proposition 8 that also the converse direction holds. In other words, that we can obtain predicative comprehension from consistency.

Lemma 12 $\mathsf{S}^1_2 + \mathsf{Con}(U_{\mathsf{V}})$ *interprets* $\mathsf{PCS}[U]$.

Proof We first define the domain(s) of our interpretation $\mathfrak{I}$. The domain of quantifiers ranging over variables of sort $\mathfrak{o}$, $\delta_{\mathfrak{I}}$, is just $\delta_{\mathfrak{F}}$ (cf. Proposition 8), that is the domain of Henkin constant in a $\mathsf{S}^1_2 + \mathsf{Con}(U)$-definable initial segment of the natural numbers. The domain of class variables, with $\mathcal{L}^c_U$ denoting again the extension of $\mathcal{L}_U$ with a countable set of Henkin constants, is

$$\delta^{\mathfrak{c}}_{\mathfrak{I}} := \{x \mid x \in \mathsf{Fml}^1_{\mathcal{L}^c_U} \cap \mathcal{I} \text{ and the free variable of } x \text{ is of sort } \mathfrak{o}\}$$

We recall that $\mathcal{I}$ is the cut defined in the proof of Proposition 8. In the official definition of the language of $\mathsf{PCS}[U]$, Xx reads $\mathsf{A}(X, x)$. Therefore we translate:

$$(\mathsf{A}(X, x))^{\mathfrak{I}} :\leftrightarrow \mathsf{S}(\ulcorner\varphi\urcorner(x/v)), \qquad\qquad \text{if } X \text{ is mapped into } \varphi(v)$$

Unsurprisingly, the clause just states that Xx is translated as 'φ is satisfied by x'. We call φ_X the element of $\delta^{\mathfrak{c}}_{\mathfrak{I}}$ associated to X by the translation.

Crucially, $\mathsf{S}^1_2 + \mathsf{Con}(U)$ proves, by meta-induction on the complexity of $\Phi(\vec{Y})$,[12] and by the crucial contribution of Proposition 8(d):

$$(\forall \vec{Y} \in \delta^{\mathfrak{c}}_{\mathfrak{I}})\big((\Phi(\vec{Y}))^{\mathfrak{I}} \leftrightarrow \mathsf{S}(\ulcorner\Phi(\vec{\varphi_Y})\urcorner)\big) \tag{14.31}$$

The base case is basically given by the definition of the translation. The other cases are easy by employing the properties of S.

To verify that $\mathfrak{I}$ is indeed a relative interpretation, we prove in $\mathsf{S}^1_2 + \mathsf{Con}(U)$ the translation of the axioms of $\mathsf{PCS}[U]$: in particular (i) axioms of the form $\forall \vec{Y}\ \Phi(\vec{Y})$ and (ii) all instances of comprehension.

(i) for $\Phi(\vec{X})$ with $\vec{X}$ arbitrary, by Lemma 8 we know that $\mathsf{Pr}_{U_{\mathsf{V}}}(\ulcorner\Phi(\vec{\varphi_X})\urcorner)$. Thus by reflection (Lemma 8(d)), $\mathsf{S}(\ulcorner\Phi(\vec{\varphi_X})\urcorner)$. By (14.31), $(\Phi(\vec{X}))^{\mathfrak{I}}$. By the properties of interpretations,, since $\vec{X}$ is arbitrary, $(\forall \vec{X}\Phi(\vec{X}))^{\mathfrak{I}}$.

Ad (ii), to any $\varphi(u, \vec{Z})$ with no bound set variables, we associate to it $\varphi_X(v_i, \vec{\varphi_Z})$ by the definition of $\mathfrak{I}$. Since φ is a standard formula, we know $\varphi_X(v_i, \vec{\varphi_Z})$ will be in $\delta^{\mathfrak{c}}_{\mathfrak{I}}$. But $\varphi_X(u, \vec{\varphi_Z})$ is just $(Xu)^{\mathfrak{I}}$, and it is equivalent to $\mathsf{S}(\varphi_X(u, \vec{\varphi_Z}))$ by the properties of $\mathfrak{I}$. By (14.31), we conclude $(\varphi(u, \vec{Z}))^{\mathfrak{I}}$. This shows that, given any $\varphi(u, \vec{Z})$, the translation of the comprehension axiom of $\mathsf{PCS}[U]$ is satisfied. □

Thus, by Lemma 3,

Corollary 9 *Let U be as above.* $\mathsf{Q} + \mathsf{Con}(U_{\mathsf{V}})$ *interprets* $\mathsf{PCS}[U]$.

[12] $\Phi(\vec{Y})$ does not contain any $\mathfrak{c}$-quantifiers.

14.4.5 Truth and Comprehension, via Consistency

Finally we can combine the claims introduced in the last two sections.

We start with predicative comprehension. By Lemmas 11 and 12, we have

Corollary 10 *Let U_{V} be sequential. Then $\mathsf{PCS}[U]$ is mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U_{\mathsf{V}})$.*

Corollary 11 *If U is finitely axiomatized and sequential, then $\mathsf{PC}[U]$ is mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U)$.*

Proof If U is finitely axiomatized, we don't need to resort to the schematic version of U to prove Lemmas 11 and 12. □

Now we move to Tarskian and positive typed truth. By Proposition 9, Corollaries 7, and 5,

Proposition 10 *(i) Let U be finitely axiomatized, then $\mathsf{T}[U]$ and $\mathsf{Tp}[U]$ are mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U)$;*
(ii) Let U be arbitrary, then $\mathsf{T}^{+}[U]$ and $\mathsf{Tp}^{+}[U]$ are mutually interpretable with $\mathsf{Q} + \mathsf{Con}(U)$.

Predicative comprehension and Tarskian truth, via consistency, enjoy a mutual reduction for a wide range of choices of the object theory:

Proposition 11 *(i) Let U be finitely axiomatized and sequential. Then $\mathsf{PC}[U]$ is mutually interpretable with $\mathsf{T}[U]$ and $\mathsf{Tp}[U]$;*
(ii) Let U_{V} be sequential. Then $\mathsf{PCS}[U]$ is mutually interpretable with $\mathsf{T}^{+}[U]$ and $\mathsf{Tp}^{+}[U]$.

14.5 Conclusion

Proof-theoretic reductions, such as the relative interpretation of PA in ZF, have often been considered as examples of ontological reductions—of natural numbers to sets, in the case at issue. Ideally, we would require a reduction of ACA to CT to replace our commitment to arithmetically definable sets with a semantic commitment to a Tarskian, compositional truth predicate. It might be thought, for instance, that any plausible criterion of ontological reduction of what is implicit in the acceptance of T to what is implicit in the acceptance of W should at least require T to be relatively interpretable in W. In such scenario, the reductions investigated in this work—and summarized in Sect. 14.4.5—seem to offer the possibility of freely moving from one's ontological commitment to subsets of the domain of the object theory to ideological commitment to concepts giving more structure to but not enlarging the original domain, such as truth and satisfaction.

These claims may be contrasted by reflecting on the combinatorial nature of a relative interpretation, which appears to be primarily a syntactical reduction: there seems to be no need to resort to a world-language relation to make sense of the interpretation of ACA in CT, for instance. It might simply be taken to be no more than a complex relation between syntactically individuated entities, that is languages and theories. The significance of the results presented in this work, however, is largely independent from one's stance on the relationships between proof-theoretic and ontological reductions. Even from a more neutral stance our study seems to reinforce the belief that set-existence principles and principles governing primitive predicates for truth and satisfaction are deeply intertwined and that a general criterion of theory choice—once we accept the base theory U, how much comprehension should we add to it? Which truth axioms for it are preferable?—should consider them as interdependent. The analysis of the operations on theories $\mathsf{T}^+[\cdot]$, $\mathsf{Tp}^+[\cdot]$ and $\mathsf{PCS}[\cdot]$ in Sect. 14.4 generalizes in fact the standard approach surveyed in Sect. 14.3 in several respects and strengthens our conviction that predicative comprehension and suitable truth axioms are indeed parallel logico-mathematical devices.

One might in fact suspect that the reductions between ACA and CT rely on principles that have not much to do with truth and predicative comprehension: by considering for instance ACA_0 and $\mathsf{CT}\restriction$, resulting from ACA and CT by restricting induction, the symmetry between truth and predicative comprehension suggested by the case with full induction is lost. ACA_0 is not interpretable in $\mathsf{CT}\restriction$. Again part of the reason for this lies with the reflexivity of PA, which is a peculiarity of the base theory and has nothing to do with truth or with predicative comprehension. $\mathsf{T}^+[\cdot]$ and $\mathsf{PCS}[\cdot]$, in this respect, fare reasonably better. Their construction is rooted in the idea that $\mathsf{T}^+[U]$ genuinely applies a syntactico/truth-theoretic package to the base theory, and that $\mathsf{PCS}[U]$ generalizes to arbitrary theories the method for obtaining ACA_0 from PA: this approach tries to minimize and control peculiar features of the object theory, such as the reflexivity of PA, for instance. Furthermore, $\mathsf{T}^+[\mathsf{PA}]$ and $\mathsf{PCS}[\mathsf{PA}]$ are comparable, as Proposition 11 shows, whereas a similar characterization for $\mathsf{CT}\restriction +$ 'all axioms of PA are true' still seems to be missing.[13] This suggests that the operations on the theories considered here offer a more uniform, smoother comparison of semantic and set existence assumptions.[14]

Another attractive feature of the proposed results is that they represent different but equivalent ways to uncover the 'lower bound' of our implicit commitment to the acceptance of a base theory U. Corollaries 10-11 and Propositions 10-11 tell us that the application of $\mathsf{T}^+[\cdot]$, $\mathsf{Tp}^+[\cdot]$ and $\mathsf{PCS}[\cdot]$ correspond to an intensional consistency statement for U, which is not reducible—neither provable nor interpretable—to U by the incompleteness phenomena. Considered in this instrumental fashion, positive, typed truth and Tarskian truth on one side and predicative comprehension on the other

[13] In particular, it is known that $\mathsf{CT}\restriction +$ 'all axioms of PA are true' interprets ACA_0, but the converse claim seems to be an open problem.

[14] As we have seen, more generality requires different means of reduction: conservativeness for suitable class of formulas (e.g. Π^0_2 or formulas of the base language), for instance, becomes a trivial requirement as all the theories resulting from the application of our functors will be conservative over the base theory. Relative interpretability becomes the preferred means of reduction.

represent equivalent routes for making explicit our commitment to U via the assertion of its consistency.

Further technical and philosophical developments are of course possible. On the philosophical side, a unified account of theory choice for systems of truth is much needed, given the explosion the research on axiomatic truth has had in recent years. From the technical standpoint, an obvious extension of the work would have to take into account hierarchies of applications of the operators: the difficulty of the hierarchical approach, in this context, seems to be represented by the fact that methods for a hierarchical analysis of $\mathsf{T}[\cdot]$, for instance, are foreseeable only when a single object theory is fixed. This would lead to a weakening of the approach that draws much of its strength from its general applicability. At any rate this is only tentative and we defer a treatment of these extensions to forthcoming works.

References

Buss, S. (1986). *Bounded arithmetic*. Naples: Bibliopolis.

Buss, S. (Ed.). (1998). *Handbook of proof theory*. Amsterdam: Elsevier.

Cantini, A. (1989). Notes on formal theories of truth. *Archives for Mathematical Logic*, *35*, 97–130.

Enayat, A., & Visser, A. (2014). New constructions of satisfaction classes. *Logic preprint group series*.

Enderton, H. (2001). *A mathematical introduction to logic*. New York: Harcourt Academic Press.

Feferman, S. (1960). Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, *49*, 35–91.

Feferman, S. (1964). Systems of predicative analysis. *Journal of Symbolic Logic*, *27*, 1–30.

Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic*, *56*, 1–49.

Fischer, M. (2009). Minimal truth and interpretability. *The Review of Symbolic Logic*, *4*(2), 799–815.

Fujimoto, K. (2010). Relative truth definability of axiomatic truth theories. *The Bulletin of Symbolic Logic*, *16*(3), 305–344.

Hájek, P., & Pudlák, P. (1993). *Metamathematics of first-order arithmetic*. Berlin: Springer.

Halbach, V. (1999). Conservative theories of classical truth. *Studia Logica*, *62*, 353–370.

Halbach, V. (2009). Reducing compositional to disquotational truth. *Review of Symbolic Logic*, *2*, 786–798.

Halbach, V. (2014). *Axiomatic theories of truth* (revised ed.). Cambridge: Cambridge University Press.

Heck, R. (2015). Consistency and the theory of truth. *The Review of Symbolic Logic, 8*(3), 424–466.

Lachlan, A. (1981). Full satisfaction classes and recursive saturation. *Canadian Mathemtical Bulletin*, *24*, 295–297.

Leigh, G. (2014). Conservativity for theories of compositional truth via cut-elimination. *The Journal of Symbolic Logic, 80*(3), 845–865.

Moschovakis, Y. (1974). *Elementary induction in abstract structures*. New York: American Elsevier.

Nicolai, C. (2015a). A note on typed truth and consistency assertions. *The Journal of Philosophical Logic*. doi:10.1007/s10992-015-9366-6

Nicolai, C. (2015b). Deflationary truth and the ontology of expressions. *Synthese*. doi:10.1007/s11229-015-0729-x

Pohlers, W. (2009). *Proof theory. The first step into impredicativity*. Berlin: Springer.

Takeuti, G. (1987). *Proof theory*. Amsterdam: North-Holland.

Vaught, R. A. (1967). Axiomatizability by a schema. *The Journal of Symbolic Logic*, *32*, 473–479.
Visser, A. (1991). The formalization of interpretability. *Studia Logica*, *50*(1), 81–106.
Visser, A. (2009). The predicative Frege hierarchy. *Annals of Pure and Applied Logic*, *160*, 129–153.

Chapter 15
A Critical Overview of the Most Recent Logics of Grounding

Francesca Poggiolesi

Abstract In this paper our aim is twofold: on the one hand, to present in a clear and faithful way two recent contributions to the logic of grounding, namely Correia (2014), and Fine (2012a); on the other hand, to argue that some of the formal principles describing the notion of grounding proposed by these logics need to be changed and improved.

Keywords Grounding · Logic · Proof theory

15.1 Introduction

The last decade has witnessed an increasing interest for the concept of grounding. Grounding is either described as a "special sort of non-causal priority", see Correia and Schnieder (2012) or as an "objective relation which is explanatory in nature", see Correia (2014). Grounding is typically conveyed by the linguistic expression 'because', but other expressions like 'in virtue of' and 'due to' can also serve the purpose. Some examples of grounding sentences are the following:

- Albert did not go to school today because he was sick;
- John is tall and thin because John is tall and John is thin;
- these two apples resemble to each other because they have the same shape and the same color.

As is clear even at the first glance, these three sentences share a common structure: each of them contains the expression *because* and each of them can be divided into an antecedent, i.e. what comes after the *because* ("Albert was sick", "John is tall and John is thin" and "the apples have the same shape and the same color"), and a

F. Poggiolesi (✉)
IHPST - Institut d'Histoire et de Philosophie, des Sciences et des Techniques, Université Paris 1 Panthéon-Sorbonne, CNRS, ENS, UMR 8590,Paris, France
e-mail: poggiolesi@gmail.com

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_15

consequent, i.e. what comes before the *because* ("Albert did not go to school", "John is tall and thin", "these two apples resemble to each other", respectively). In each case we can say that the consequent is *determined*, or *explained* or *accounted by* the antecedent. In other worlds, in each of the sentences listed above, the antecedent constitutes the *reason why*, or the *ground* of the consequent.

Most of the work on grounding is metaphysical (e.g. see Betti 2010; Daily 2012; Schaffer 2009). Beside this kind of study, two other types of research, that are equally worthy of attention, has been carried out: on the one hand, there exists a rising number of papers dedicated to the history of the notion of grounding, in particular to Bernard Bolzano, a precursor in the study of this concept (e.g. see Betti 2010; Tatzel 2002). On the other hand, some scholars have directed their interest to the creation of different logics of grounding, namely Batchelor (2010), Correia (2010, 2014), Fine (2012a, b), Schnieder (2011). These works have undoubted value: they have indeed opened up the new and important topic of what the formal properties and the basic principles of the concept of grounding are.

In this paper our aim is twofold: on the one hand, to focus on the two most recent works in the logic of grounding, namely on Correia (2014) and Fine (2012a), and to summarize them in a clear and faithful way; on the other hand, to concentrate on the formal principles describing the notion of grounding proposed by these logics and to argue that some of these need to be changed and improved. The paper is organized as follows. Section 15.2 will serve to introduce certain important distinctions concerning the concept of grounding which will prove useful in what follows; Sect. 15.3 will be used to illustrate the conception of grounding which is adopted in this paper for evaluating the two most recent logics of grounding. In Sect. 15.4 we will present the logic of Fine (2012a), while in Sect. 15.5 we will discuss that of Correia (2010). Each of the Sects. 15.6–15.8 will be dedicated to a different problem linked to the aforementioned logics of grounding and to its discussion. Finally, in Sect. 15.9 we will draw some conclusions.

15.2 Some Standard Distinctions Concerning the Notion of Grounding

We will use this section to introduce three familiar distinctions concerning the concept of grounding: the first is that between full and partial grounding, the second is the distinction between immediate and mediate grounding; the third is the distinction between strict and weak grounding. These distinctions have been defined and used by different authors (e.g. see Bolzano 1973; Correia and Schnieder 2012); here we introduce them in the form proposed by Fine (2012a), which seems to us the form most commonly accepted in the contemporary literature.

Let us start with the distinction between full and partial grounding. A set of truths M is a *full* ground of the truth C if M is sufficient to explain the truth C; A is a *partial* ground of C if A on its own, or with some other truths, is a full ground of C.

> Thus, given that A, B is a full ground for $A \wedge B$, each of A and B will be a partial ground for $A \wedge B$. Each will be relevant to the grounding of $A \wedge B$, even though neither may be sufficient on its own. (Fine 2012a, p. 50)

Let us now pass to the second distinction, the one between immediate and mediate grounding. A is an *immediate* ground of C if C may be obtained from A by means of a single grounding step; A is a *mediate* ground of C if C is obtained from A by appropriately chaining immediate grounding steps. Thus, while $A \wedge B$ is immediately grounded in A and B, $(A \wedge B) \wedge C$ is only mediately grounded in A and B and C.

Before passing to the third and last distinction concerning the grounding concept, let us underline the following point, which will be important later on. We have identified four different types of ground: full and immediate, full and mediate, partial and immediate and partial and mediate. According to the way they have been defined, these types of ground are not pairwise disjoint, i.e. given a set of truths M and a truth C there might be more than one type of grounding relation that intervenes between M and C. This is clearly demonstrated by two simple examples. Consider first of all the two truths A and $A \vee B$; according to the definitions above, A is both the full and immediate, but also the partial and immediate ground of $A \vee B$. A is the immediate ground of $A \vee B$ because $A \vee B$ can be obtained from A thanks to an unique grounding step. A is both a full and a partial ground of $A \vee B$ because A is sufficient, but also relevant, to guarantee the truth of $A \vee B$.

Now consider the two truths A and $A \vee (A \vee A)$. According to the definitions above, A and $A \vee (A \vee A)$ enjoy all four types of grounding relation. A is at the same time the full and the partial ground of $A \vee (A \vee A)$ for reasons analogous to the ones mentioned for A and $A \vee B$. As for A being both immediate and mediate ground of $A \vee (A \vee A)$, we can consider the following explanation

> The truth that A, for example, is a ground for $A \vee (A \vee A)$. It is furthermore an immediate ground for $A \vee (A \vee A)$ since A in its capacity as a left disjunct, so to speak, is not a mediated ground for $A \vee (A \vee A)$. However, A is an immediate ground for $A \vee A$ and $A \vee A$ is an immediate ground for $A \vee (A \vee A)$; and so A also stand in a mediated relationship of ground to $A \vee (A \vee A)$. (Fine 2012a, p. 51)

Let us now pass to the third and last distinction concerning grounding that is of interest in this paper. A relation of grounding is *strict* when it does not allow a truth to partially ground itself; a relation of grounding is *weak* when it does. Traditionally (e.g. Bolzano 1996) only strict grounding is taken into account; the relation of weak grounding has been introduced in the recent literature.

In Sects. 15.4 and 15.5 we will present the logics introduced in Fine (2012a) and Correia (2014), respectively. Instead of considering these logics in the framework of the first-order language, which is the one used by their authors, we will only focus on the framework of the propositional language; moreover, we will restrict the logics of Fine and Correia to the only notion of (weak or strict) full and immediate grounding. There are several reasons for adopting these restrictions. First of all, we do not want the paper to be burdened with too many technical notions and thus we believe that some simplifications are appropriate. Secondly, both the domain of propositional language and the concept of (weak or strict) full and immediate grounding are already rich and fertile enough to inspire reflexions. Thirdly, we privilege the concept of (weak or strict) full and immediate grounding over the others since, as underlined by Bolzano (1973), Tatzel (2002), it is central and, by dealing with it, one really touches the hearth of the matter.

Note that no choice is made with regard to the distinction between strict and weak grounding. Despite the fact that some scholars, like de Rosset (2013), have criticized the notion of weak grounding and we are quite sympathetic to this position, we need both the notions of strict and weak grounding in order to introduce some salient features of the logics of Fine and Correia (e.g. without the weak notion of grounding the elimination rules of Fine's logic could not be introduced.) This is the reason why in this case we have not adopted any restriction.

15.3 Grounding as a Proof-Theoretic Notion

As already emphasized in the introduction, grounding is a deep and complex concept that is and has been studied from several different perspectives, e.g. metaphysical, historical, logical. In this paper we will take a point of view which has recently received a renovated attention (e.g. see Rumberg 2013; Tatzel 2001) and that could be classified as proof-theoretical: the logics of Fine and Correia will thus be evaluated against such a background. We will use the rest of the section to illustrate the proof-theoretic conception of grounding and justify its use in the critical analysis of the logics that can be found in Correia (2014) and Fine (2012a).

In 1935 the logician Gherard Gentzen introduced *the calculus of natural deduction*. The calculus of natural deduction is a formal system, just like a Hilbert system, where the focus does not go on axioms but on rules. The rules of natural deduction are typically many and can be divided into introduction and elimination rules: each of the former introduces a different logical connective, while each of the latter eliminates a different logical connective. By means of these rules, one can construct derivations that have the form of trees whose leaves correspond to assumptions and whose root corresponds to the conclusion (for a detailed introduction to the calculus of natural deduction see Troelstra and Schwichtenberg 1996).

Many interesting theorems concerning the calculus of natural deduction can be proved. First of all it can (and should) be shown that a calculus of natural deduction is sound and complete with respect to a certain semantics or a corresponding Hilbert system. Secondly, given a calculus of natural deduction, it can be proved that several rules are admissible and others are derivable in it. Let us remind the reader (since it will prove useful in what follows) that a rule $\mathcal{R}$, that does not belong to a calculus of natural deduction **N**, is said to be *admissible* in **N** if, whenever there exists a derivation in **N** of the premise(s) of $\mathcal{R}$, then there also exists a derivation in **N** of the conclusion of $\mathcal{R}$, that does not contain any application of $\mathcal{R}$. A rule $\mathcal{R}$, that does not belong to a calculus of natural deduction **N**, is said to be *derivable* in **N** if a derivation from the premisses of $\mathcal{R}$ to the conclusion of $\mathcal{R}$ can be constructed in **N**.

Thirdly and finally, given a calculus of natural deduction, the famous *normalization* theorem can (and should) be established. This theorem can be roughly explained in the following way. Consider a derivation that contains at least one formula that is neither one of its premisses, nor its conclusion, nor a subformula of its conclusion; such a derivation contains a *redundancy*, or, in other words, it is a *non-analytic derivation*. The normalization theorem ensures that any derivation that can be constructed in a calculus of natural deduction is either analytic or can be effectively transformed into an analytic one. The importance of the normalization theorem and of analytic derivations has a long and venerable history that starts with Aristotle (e.g. see Paoli 1991), pass through Gentzen (1935) and continues today (e.g. see Avron 1991; Indrezejczak 1997; Poggiolesi 2012; Prawitz 1965): the normalization theorem is by now a cornerstone result of proof theory and analytic derivations represent a paradigm that can hardly be ignored.

Let us now move back to the concept of grounding and consider the link between this concept and the calculus of natural deduction that has so far captured our attention. For many illustrious scholars (e.g. Betti 2010; Bolzano 1996; Casari 1987; Sebestik 1992), the relationship is simple: grounding is nothing but a special sort of derivability; more precisely, grounding proofs should be seen as particular types of derivations, which reveal ontological hierarchies where truths have been arranged. In this sense grounding is a proof-theoretic notion. If we embrace such a view, as we do in the present paper, there are consequences that cannot be neglected. First of all, the formalization of the grounding notion will need to be accomplished by means of a calculus comparable to the one of natural deduction; moreover, such a calculus for grounding will need to satisfy and respect many (if not all) of the properties that have been put forward for the standard calculus of natural deduction. Amongst these, the analyticity property, whose centrality has been emphasized above, will need to keep on playing a special role: grounding chains, as peculiar derivations, will need to be analytic. It is interesting to notice that Bernard Bolzano, the founding father of the theory of grounding, stressed the importance of the analyticity property for grounding chains for reasons independent from the recent developments in contemporary proof theory (see Paoli 1991).

Correia (2014) and Fine (2012a) present a calculus, as well as a semantics, for the notion of grounding. While Correia proves that his calculus is sound and complete with respect to the semantics that he proposes, the semantics of Fine (2012a) is not adequate for his system (Correia and Schnieder 2012, p. 19). In the next sections, we will only concentrate on the calculi introduced by Fine and Correia. As it seems natural to do, we will evaluate these calculi from a proof-theoretic perspective, that is from the perspective that we have presented in this section.

15.4 Fine's Logic

We will use this section to introduce Fine (2012a) logics for the concepts of strict and weak, full and immediate grounding.

Definition 15.4.1 Let $\mathcal{L}$ be the language composed by:

- propositional atoms: p, q, r, ...
- classical connectives: $\neg$, $\wedge$, $\vee$
- $>$ (strict, full and immediate grounding), $\geqslant$ (weak, full and immediate grounding)

As metalinguistic symbols we will use the comma, the semi-colon and the parentheses.

Definition 15.4.2 The propositional formulas of the language $\mathcal{L}$ are defined by means of the following rule:

$$p|\neg A|A \wedge B|A \vee B$$

The grounding formulas of the language $\mathcal{L}$ are of one of these two types:

$$M > A$$

$$M \geqslant A$$

where M is a set of propositional formulas and A is a propositional formula.

The calculus **FG** (for Fine on grounding) is composed by axioms, see Fig. 15.1, logical rules, see Fig. 15.2, and structural rules, see Fig. 15.3; we will analyse each of these components, starting from structural rules. These are three: the rule SW and the rules ID and NID. The rule SW tells us that from the fact that a relation of strict grounding holds between a set of truths M and a truth A we can infer that a relation of weak grounding holds amongst these truths. The rules ID and NID exemplify the difference between strict and weak grounding: the former is irreflexive, the latter is reflexive. Note that in the original calculus created by Fine the structural rules are much more numerous than the ones that appear in Fig. 15.3: they indeed serve

$$A > \neg\neg A$$

$$A, B > A \wedge B \qquad A > A \vee B \qquad B > A \vee B \qquad A, B > A \vee B$$

$$\neg A > \neg(A \wedge B) \qquad \neg B > \neg(A \wedge B) \qquad \neg A, \neg B > \neg(A \wedge B) \qquad \neg A, \neg B > \neg(A \vee B)$$

Fig. 15.1 Axioms of the calculus **FG**

$$\frac{M > \neg\neg A}{M \geqslant A}\ \neg E$$

$$\frac{M > A \wedge B}{M \geqslant \{A, B\}}\ \wedge E \qquad \frac{M > A \vee B}{M \geqslant \{A, B\}; M \geqslant B; M \geqslant A}\ \vee E$$

$$\frac{M > \neg(A \wedge B)}{M \geqslant \{\neg A, \neg B\}; M \geqslant \neg B; M \geqslant \neg A}\ (\neg\wedge)E \qquad \frac{M > \neg(A \wedge B)}{M \geqslant \{\neg A, \neg B\}}\ (\neg\vee)E$$

Fig. 15.2 Logical rules of the calculus **FG**

Fig. 15.3 Structural rules of the calculus **FG**

$$\frac{M > C}{M \geqslant C}\ SW$$

$$\frac{}{A \geqslant A}\ ID \qquad \frac{M, A > A}{\perp}\ NID$$

to describe the relationships amongst the several types of grounding (partial and immediate, partial and mediate ...) that have been mentioned in Sect. 15.2 but are not discussed here.

Let us now move to the logical rules of Fig. 15.2. These rules are called by Fine *elimination rules*, just as the elimination rules of the classical natural deduction calculus. Note however an important (from a proof-theoretical point of view, see Indrezejczak 1997; Poggiolesi 2010) difference between the classical rules and the ones introduced by Fine: while each of the former simply eliminates one classical connective a time, each of latter eliminates one or two classical connectives a time, plus it makes us move from strict to weak grounding. Let us emphasize that in the rules $\vee E$ and $(\neg\wedge)E$ the (meta-linguistic symbol) semi-colon is used: it indeed serves to indicate the disjunctive character of the conclusion.

Let us finally comment on the syntactic objects of Fig. 15.1. Whilst Fine call these objects *introduction rules*, we prefer to call them axioms since none of them involve an inferential step. These axioms describe the strict, full and immediate grounds of the classical connectives; let us note that negation, differently from conjunction and disjunction, is treated only in its interaction with other connectives.

We would like to end this section with two further remarks. The first concerns the notion of derivation in the calculus **FG**. It is not clear how to adapt the standard definition of derivation to the calculus **FG** because of the presence of the (meta-linguistic symbol) semi-colon, for which no indication about use is given.[1]

[1]The use of the semi-colon, whose interpretation is in disjunctive terms, could lead one to think of *hypersequents* (e.g. see Poggiolesi 2008, 2013). Despite the analogous interpretation, the calculus of Fine and hypersequent calculi are different. In hypersequent calculi not only do we have external structural rules that tell us how to deal with the semi-colon, but also the logical rules are general

The second remark concerns the rule of *amalgamtion*, which has the following form:

$$\frac{M_1 > C, M_2 > C, \dots}{M_1, M_2, \dots > C}\ am$$

The rule basically says that the strict grounds of a given truth can be amalgamated or combined into a single ground.

> It is not usual to include this rule among the rules for ground. But the plausibility of the rules from which it can be derived provides a strong argument for its adoption; and I doubt there is a simple and natural account of the logic of ground that can do without it. (Fine 2012a, p. 57)

It follows from amalgamation that there always is a maximum full and immediate ground for a grounded truth C: if $M > A$, then there is a N such that (i) $N > A$, and (ii) $P \subseteq N$, for all P such that $P > A$. Note that the same holds for the notion of full and mediate ground.

It is easy to see that the rule of amalgamation is admissible in the calculus **FG**. Indeed its premisses can only be the axioms of **FG** and the only cases where these axioms have the same conclusion are the disjunctive and the negation of conjunction ones; the use of the axioms $A, B > A \vee B$ and $\neg A, \neg B > \neg(A \wedge B)$, respectively, makes the rule straightforwardly admissible.

15.5 Correia's Logic

In this section we introduce Correia (2014) logics for the concepts of strict and weak full and immediate grounding. We underline that Correia (2014) never mentions the distinction between immediate and mediate grounding and his results concern mediate grounding. Nevertheless, it is quite trivial to see how to adapt his logic and his results to the only case of immediate grounding.

Definition 15.5.1 Let $\mathcal{L}'$ be the language composed by: propositional atoms: p, q, r, . . . ; the classical connectives: $\neg$, $\wedge$, $\vee$, and the parentheses: (,). The formulas of the language $\mathcal{L}$ are defined standardly. $M, N, \dots$ will denote sets of formulas.

The calculus **CG** (for Correia on grounding) is composed by the rules in Fig. 15.4, which basically correspond to the axioms of the calculus **FG**.

Definition 15.5.2 We say that M strictly, fully and immediately ground A, in symbols $M \triangleright A$, if, and only if, M and A can be connected by means of one of the rules of the calculus **CG**.

(Footnote 1 continued)
enough to cover the whole hypersequent object. None of these features is present in **FG** and this is the reason why it is to hard to figure out how to adapt the standard notion of derivation to this calculus.

Fig. 15.4 Logical rules of the calculus **CG**

$$\frac{A}{\neg\neg A}\ {\scriptstyle \neg\neg I}$$

$$\frac{A, B}{A \wedge B}\ {\scriptstyle \wedge I} \qquad \frac{A}{A \vee B}\ {\scriptstyle \vee I} \qquad \frac{B}{A \vee B}\ {\scriptstyle \vee I}$$

$$\frac{\neg A}{\neg(A \wedge B)}\ {\scriptstyle (\neg\wedge)I} \qquad \frac{\neg B}{\neg(A \wedge B)}\ {\scriptstyle (\neg\wedge)I} \qquad \frac{\neg A, \neg B}{\neg(A \vee B)}\ {\scriptstyle (\neg\vee)I}$$

Definition 15.5.3 We say that M weakly, fully and immediately ground A, in symbols $M \blacktriangleright A$, if, and only if, $A \in M$ or, for some $N \subseteq M$, $N \rhd A$.

As Correia (2014), p. 8 himself remarks, his notion of weak ground is significantly different from the one that we saw in Fine: while his notion obeys weakening, Fine's does not. It has to be said that doubts may be expressed as to whether a notion of grounding should satisfy weakening: indeed it seems intrinsic to the concept of grounding that all grounds are relevant to the conclusion, while weakening notoriously violates this condition.

Let us list some properties of $\rhd$ and $\blacktriangleright$.

Proposition 15.5.4 *For any M and any A,*

- *if $M \rhd A$, then $M \neq \emptyset$,*
- *if $M \rhd A$, then A is not an atom, nor the negation of an atom,*
- *It is not the case that $M, A \rhd A$,*
- *if $M \rhd A$, then the complexity of A is strictly greater than the complexity of any member of M, following the standard notion of complexity of a formula,*[2] *e.g. see Troelstra and Schwichtenberg (1996).*

Proposition 15.5.5 *For any M and any A,*

- *if $M \blacktriangleright A$, then $M \neq \emptyset$,*
- *if $M \blacktriangleright A$ and $A \notin M$, then A is not an atom, nor the negation of an atom,*
- *$M, A \blacktriangleright A$,*
- *if $M \blacktriangleright A$, then $M, N \blacktriangleright A$.*

Let us now consider the calculus **GC**$'$; **GC**$'$ is obtained from **GC** by adding the following two rules:

$$\frac{A, B}{A \vee B}\ {\scriptstyle \vee I} \qquad \frac{\neg A, \neg B}{\neg(A \wedge B)}\ {\scriptstyle (\neg\wedge)I}$$

Definition 15.5.6 $M \rhd' A$, if, and only if, M and A can be connected by means of one of the rules of the calculus **CG**$'$.

[2]Let us underline that Correia uses a notion of complexity which is slightly different from the one that can be standardly found in the literature. Nevertheless Claim 4 of Proposition 15.5.4 holds for both notions.

Definition 15.5.7 $M \blacktriangleright' A$, if, and only if, $A \in M$ or, for some $N \subseteq M$, $N \rhd' A$.

Correia works with the aforementioned notions $\rhd$ and $\blacktriangleright$; nevertheless he considers $\rhd'$ and $\blacktriangleright'$ as two serious alternative logics. While $\rhd$ and $\rhd'$ clearly do not have the same extension, $\blacktriangleright$ and $\blacktriangleright'$ are equivalent.

In the last section we have underlined the importance of the amalgamation rules in Fine's logic. Such a rule turns out to be admissible in the calculus **GC**$'$ with respect to both notions $\rhd'$ and $\blacktriangleright'$. The proof of the admissibility of the rule is analogous to the one that we have sketched in the case of Fine's logics (see Sect. 15.4).

15.6 Negative Truths

Only a few logics have been proposed for the concept of grounding; in the last sections we have introduced and illustrated two of the main ones, namely those of Fine and Correia. In this section and following two, we will analyze these logics in detail. Our analysis will not be technical in nature, i.e. we will not discuss the technical virtues and defects of the calculi **FG** and **CG**; rather we will focus on the formal principles governing the concept of full and immediate grounding that these calculi propose and we will evaluate them from a proof-theoretic perspective. We will argue that some of these principles need to be changed, while other principles need to be added. More precisely, while in this section we will discuss principles concerning grounding and negation, in the next section we will discuss principles concerning grounding and disjunction; finally in Sect. 15.8 we will discuss principles concerning grounding and truths that are equivalent under applications of associativity and commutativity of conjunction and disjunction.

Let us start by focussing on the axiom $\neg A, \neg B > \neg(A \vee B)$ of **FG** and the rule $(\neg\vee)I$ of **CG**. The axiom and the rule basically say the same thing: the strict, full and immediate grounds of any formula of the form $\neg(A \vee B)$ are $\neg A$ and $\neg B$. Let us confront this statement with some concrete examples. Let us consider the truth "it is not the case that Mary is tall or thin" and let us ask ourselves what the strict, full and immediate grounds of this truth are. The natural answer to such a question is "Mary is not tall" and "Mary is not thin": both these truths seem to be the (full and immediate) reasons of the truth "it is not the case that Mary is tall or thin"; thus the answer agrees with the logics **FG** and **CG**.

Let us now take another example. Consider the truth "it is not the case that Mary is not tall or not thin": what are its strict, full and immediate grounds? A little reflection suffices to realize that in this case the natural answer is "Mary is tall" and "Mary is thin": "Mary is tall" and "Mary is thin" seem indeed to be the (full and immediate) reasons why "it is not the case that Mary is not tall or not thin" is true. In this case the answer differs from that given by the logics **FG** and **CG**: according to their axioms and rules, the strict, full and immediate grounds of "it is not the case that Mary is not tall or not thin" are "it is not the case that Mary is not tall" and "it is not the case that Mary is not thin". So in this case intuition seems to suggest that the rules of Fine and Correia are not correct. Since intuitions might be revealing but are not always trustworthy, let us develop a more accurate analysis of the situation.

In order to be brief but precise, let us first of all use the following formalization:

- "Mary is tall" and "Mary is thin" are formalized by p and q, respectively,
- "it is not the case that Mary is not tall" and "it is not the case that Mary is not thin" are formalized by $\neg\neg p$, $\neg\neg q$, respectively,
- "it is not the case that Mary is not tall or not thin" is formalized by the formula $\neg(\neg p \vee \neg q)$.

The issue here is to establish what the strict, full and immediate grounds of $\neg(\neg p \vee \neg q)$ are: p and q or $\neg\neg p$ and $\neg\neg q$.

One good reason for choosing p and q over $\neg\neg p$, $\neg\neg q$ appeals to a question of complexity. Indeed, if we consider the standard way of measuring the complexity of a formula,[3] but also, and most importantly, if we take into account the measure of complexity put forward by Correia (2014), p and q are less complex than $\neg\neg p$ and $\neg\neg q$. Grounding is essentially an explanatory relation and, as Ockham teaches us, explanans that are less complex are to be preferred over explanans that are more complex. Thus p and q are the strict, full and immediate grounds of $\neg(\neg p \vee \neg q)$, contrary to what logics **FG** and **CG** claim.

This argument for opting for p and q over $\neg\neg p$, $\neg\neg q$ seems correct; nevertheless, in view of the distinctions that we have introduced in Sect. 15.2, one could easily reply to it as follows. p and q certainly represent the grounds for the truth $\neg(\neg p \vee \neg q)$; however they do not substitute $\neg\neg p$ and $\neg\neg q$ in the role of strict, full and immediate grounds of $\neg(\neg p \vee \neg q)$, since they are not its strict, full and immediate grounds, rather they are its strict, full and *mediate* grounds. To see this, let us consider the calculus **CG** (we could have equivalently used the calculus **FG**); in this calculus we can construct the following grounding chain, that we will call *gc*:

$$\dfrac{\dfrac{p}{\neg\neg p} \qquad \dfrac{q}{\neg\neg q}}{\neg(\neg p \vee \neg q)}$$

Such a chain evidently shows[4] that p and q are the (strict, full and) mediate grounds of $\neg(\neg p \vee \neg q)$. Thus p and q are certainly less complex than $\neg\neg p$ and $\neg\neg q$ according to two significant measures of complexity; nevertheless they are not the full and immediate grounds of $\neg(\neg p \vee \neg q)$; they are indeed its full and mediate grounds.

Let us dwell for a moment on the grounding chain *gc* that we have constructed above. Even at the first glance, this chain has the following striking property: two

[3]The complexity of a formula A, $cm(A)$, is inductively defined in the following way: $cm(p) = 0$, $cm(\neg A) = cm(A) + 1$ and $cm(A \circ B) = cm(A) + cm(B) + 1$, where $\circ = \wedge, \vee$. E.g. see Troelstra and Schwichtenberg (1996).

[4]Even if we did not formally introduce the notions of grounding chain and of full and mediate grounding in the calculus **CG** (for a detailed description, see Correia 2014), it is quite straightforward to understand them.

formulas are introduced, namely $\neg\neg p$ and $\neg\neg q$, which then disappear in the conclusion, i.e. neither one nor the other is a subformula of the formula that appears in the conclusion. In the light of what has been said in Sect. 15.3, this fact is alarming. Even if we do not have a rigorous definition of the notion of subformula for the grounding framework and thus we do not have a proper notion of analyticity for grounding chains,[5] a phenomenon of this sort cannot but be suspicious: it is the typical example of loss of analyticity in the standard calculus of natural deduction. In order to get a better grasp of the situation, we can thus proceed in the following way. We try to reconstruct the steps from p and q to $\neg\neg p$ and $\neg\neg q$ to $\neg(\neg p \vee \neg q)$ in the calculus of natural deduction. If we come up with a non-analytic derivation, then this will be a clear sign of the fact that gc is a non-analytic grounding chain.[6] We have the following derivation:

$$\dfrac{\dfrac{\cancel{\neg p \vee \neg q} \quad \dfrac{\cancel{\neg p} \quad \dfrac{\dfrac{\cancel{\neg p} \quad \mathbf{p}}{\bot}}{\neg\neg\mathbf{p}}}{\bot} \quad \dfrac{\cancel{\neg q} \quad \dfrac{\dfrac{\cancel{\neg q} \quad \mathbf{q}}{\bot}}{\neg\neg\mathbf{q}}}{\bot}}{\bot}}{\neg(\neg\mathbf{p} \vee \neg\mathbf{q})}$$

Such a derivation contains a *cut*, namely the introduction of a negation and its successive elimination; a cut is nothing but the formal counterpart of a non-analytic inferential step. Thus the derivation that leads from p and q to $\neg\neg p$ and $\neg\neg q$ to $\neg(\neg p \vee \neg q)$ in the calculus of natural deduction is non-analytic; and hence so is the chain gc that we have constructed above. It is very important to note that no derivation containing a cut arises when we reproduce in the calculus of natural deduction the grounding steps from p and q to $\neg(\neg p \vee \neg q)$, and from $\neg p$ and $\neg q$ to $\neg(p \vee q)$: these grounding steps are those that we previously classified as valid.

We can take this situation to be the formal counterpart of the intuitions that we have exposed at the beginning of the section. Given the truth "it is not the case that Mary is not tall or not thin", it seems natural to claim that its full and immediate grounds are the truths "Mary is tall" and "Mary is thin". As a matter of fact, even formally the step from "Mary is tall" and "Mary is thin" to "it is not the case that Mary is not tall or not thin" corresponds to a normal derivation in the calculus of natural deduction (and thus, most likely, to an analytic grounding step). By contrast, if we try to claim that the truths "Mary is tall" and "Mary is thin" are instead the full and mediate grounds of the truth "it is not the case that Mary is not tall or not thin", since the full and immediate grounds of this latter truth are the truths "it is not the

[5]This paper is not the place for introducing a rigorous notion of analyticity for the grounding framework since it only concerns a critical analysis of the logics of Fine and Correia. We nevertheless believe that such a task should be seriously taken into account in the studies dedicated to the concept of grounding.

[6]Under the assumption that grounding proofs are nothing but particular type of derivations, if the derivation that is behind a specific grounding chain is not analytic, then the grounding chain is not analytic too.

case that Mary is not tall" and "it is not the case that Mary is not thin", not only do we seem to be making an artificial construction, but also, at the formal level, we are compelled to accept a non-analytic grounding chain (because we have a non-normal derivation in the calculus of natural deduction).

Therefore it seems reasonable to conclude that not all instances of the axiom $\neg A, \neg B > \neg(A \vee B)$ of **FG** and the rule $(\neg\vee)I$ of **CG** are correct; in particular the only case where they work adequately is when neither in A nor in B the principal connective is a negation; the other cases need a more careful treatment. Similar conclusions hold for the axioms $\neg A, \neg B > \neg(A \wedge B)$, $\neg A > \neg(A \vee B)$, $\neg B > \neg(A \vee B)$ and all the rules $(\neg\wedge)I$.

15.7 Disjunctive Truths (and the Full-Partial Distinction)

We will use this section to point out some flaws of both the distinction between full and partial grounding, as it has been introduced in Sect. 15.2, and the axioms of **FG** concerning disjunction and grounding. As one shall argue, these criticisms are interrelated.

Before starting, let us emphasize two things. First of all, we point out that anything that will be said of the axioms of **FG** also holds for the corresponding rules of the system **CG**$'$. Secondly, we stress that the criticisms directed to the axioms of **FG** concerning disjunction and grounding also hold for those concerning negation of conjunction. For the sake of brevity, we will only consider the former and not the latter.

Let us begin by noting a significant mismatch between the full-partial distinction as introduced in Sect. 15.2 and certain axioms of **FG** that are supposed to reflect this distinction. For this, let us consider the three axioms of **FG** describing disjunction and grounding, namely $A > A \vee B$, $B > A \vee B$, $A, B > A \vee B$ and let us distinguish two ways of reading the full-partial distinction as it has been introduced in Sect. 15.2: a way that might be called *minimal* and a way that might be called *non-minimal*. According to the minimal way of reading the full-partial distinction, to give the full grounds of $A \vee B$ amounts to give exactly sufficient conditions for the truth $A \vee B$ and nothing more. In this context, while the axioms $A > A \vee B$, $B > A \vee B$ properly reflect the relation of full grounding, the third axiom does not reflect this distinction at all: A and B together are indeed more than sufficient for explaining the truth $A \vee B$. Let us then turn to the non-minimal way of reading the full-partial distinction. According to the non-minimal way of reading the full-partial distinction, to give the full grounds of $A \vee B$ may include giving even more than what is strictly sufficient for the truth $A \vee B$. In this context the three axioms $A > A \vee B$, $B > A \vee B$, $A, B > A \vee B$ all reflect the relation of full grounding. The problem in this case is rather the following: if we accept this non-minimal reading of the distinction full-partial, then we should also be willing to accept axioms of the form $A, C > A \vee B$, $B, C > A \vee B$, $A, B, C > A \vee B$ ... : each of these axioms indeed provides more than sufficient conditions for the truth $A \vee B$ and has thus the same status as the three axioms $A > A \vee B$, $B > A \vee B$, $A, B > A \vee B$. Under the non-minimal reading of the full-partial distinction, the logic for the notion of full and immediate ground should thus include not only the axioms $A > A \vee B$, $B > A \vee B$, $A, B > A \vee B$, but also axioms of the form $A, C > A \vee B$, $B, C > A \vee B$, $A, B, C > A \vee B$ .... Such a conclusion is of course unacceptable and thus the non-minimal reading of the full-partial distinction is to be rejected. We should thus stick to the minimal reading of the full-partial distinction and find a solution for the tension between such a reading and the axiom $A, B > A \vee B$. More precisely, the issue is that of understanding whether we should keep the full-partial definition (under a minimal reading) and reject the axiom, or whether, on the contrary, it is preferable to keep the axiom and reject the definition (under a minimal reading); in what follows we give an argument in favor of the second option.

For this let us consider a world where both A and B are true. If we follow the distinction full-partial, as it has been defined in Sect. 15.2, and the axioms that properly reflect it, A (or similarly B) is, by itself, the strict full and immediate ground of $A \vee B$. Is this really the case? Generally speaking, the grounding relation is indeed a relation of strong connection and authentic dependence between premisses and conclusion. This is even more true of the relation of full and immediate grounding that is, among the four relations of grounding, the most stringent and fundamental (e.g. see Tatzel 2002). Thus, we would like to claim that, in a full and immediate grounding relation, if the antecedent were not true, the consequent would not be true either. Unfortunately, this is not what happens with the full and immediate grounding relation that we have just stated: in a world where both A and B are true, the fact that A is the full and immediate ground of $A \vee B$ does not entail that, if A were not true, $A \vee B$ would not be true either, since $A \vee B$ would still be true because of B. Thus, in a world where both A and B are true, A does not seem to share such a strict connection with the truth $A \vee B$; in other words, A does not seem to be the strict, full and immediate ground of $A \vee B$. In such a situation it seems rather that both A and B are the full and immediate grounds of $A \vee B$: they both contribute, and hence explain, the truth $A \vee B$. Moreover, they stand in authentic dependency with $A \vee B$: if both A and B were not true, $A \vee B$ would not be true either. Therefore, in a world where both A and B are true, the strict, full and immediate grounding relation for the truth $A \vee B$ is properly described by the axiom of the form $A, B > A \vee B$. This seems to us as a very good reason for keeping such an axiom and, in view of its incompatibility with the full-partial distinction, for rethinking this distinction.

The second criticism that we propose in this section has to do with something that has already been mentioned in Sect. 15.2. In that section we introduced two distinctions, full-partial and immediate-mediate, that combined together identify four different types of grounding: full and immediate, full and mediate, partial and immediate and partial and mediate. The reason for introducing these four types of grounding cannot but be classificatory; in the same way as, once one has introduced the concept of natural number, then each number is either even or odd, thus, once one has introduced the concept of grounding, any grounding relation would better be of one, and only one, of the four types of grounding relation mentioned above. This is not what happens with our classification: given a set of truths M and a truth A

that stand in a grounding relation, this grounding relation can be of more than one type; actually, it can be of all four types at the same time, as is the case for A and $A \vee (A \vee A)$ (see Sect. 15.2). Such a fact should invite a serious reconsideration of our distinctions: if the limits that they describe are so wide, does this not weaken their power? Clearly, we would conceptually gain much if we could draw the distinctions between full and partial and immediate and mediate in such a way that the four types of grounding never overlap (or at least as little as possible).

Note that such a criticism of the full-partial and immediate-mediate distinctions also holds for certain axioms and rules of the logic of Fine (2012a).[7] Consider for example the following two axioms that describe the full and immediate grounds of a disjunctive truth, $A > A \vee B$ and $B > A \vee B$, and suppose that one adds the symbol $\succ$ for partial and immediate ground. The complete version of the logic of Fine (2012a) contains this rule:

$$\frac{M, A > C}{A \succ C}$$

which allows us to infer the following theorems: $A \succ A \vee B$ and $B \succ A \vee B$. Thus, in the logic of Fine it holds that A, or equivalently B, is at the same time the full and the partial, immediate grounding of the truth $A \vee B$. The same ambiguity, which has been previously pointed out for the informal full-partial and immediate-mediate distinctions, can thus now be noted for the formal principles that govern the logic introduced in Fine (2012a). This suggests that certain axioms and rules of this logic should be also carefully reconsidered.

The full-partial and immediate-mediate distinctions have been related to the notion of grounding since (at least) the XIX century. Bolzano (1996) introduced them for the first time and defined them in a way that overcomes the problems outlined in this section. Despite the fact that nowadays most philosophers have emphatically neglected Bolzano's work and embraced Fine's demarcations, the criticisms that have been put forward in this section suggest that such a choice has not necessarily been a good one. We think that a renewed attention should be dedicated to the reflexions of the great Bohemian philosopher and clearer and more precise definitions of the full-partial (and immediate-mediate) distinction should be investigated.

15.8 Associatively and Commutatively Equivalent Truths

We use this section to discuss some formal principles concerning grounding that are neglected by the logics of Fine and Correia, but that, we will argue, represent important features of the concept of grounding.

[7] This is not immediately clear from what we have presented in Sect. 15.4, since in that section we have restricted ourselves to the only case of full and immediate grounding. A quick look to Fine (2012a) will be enough to verify what we are saying.

Let us start considering the truth “Mary is tall and thin, and blond” and let us recall the strict, full and immediate grounds of this truth according to the logics **FG** and **CG**: “Mary is tall and thin” and “Mary is blond” very naturally fill this role. Let us now consider the truths “Mary is thin and tall” and “Mary is blond” and let us consider the relation between these truths and the truth “Mary is tall and thin, and blond”. Without any doubt one would like to say that “Mary is thin and tall” and “Mary is blond” are the grounds of “Mary is tall and thin, and blond”: whatever conception of ground one might have, such a conception must contain “Mary is thin and tall”, “Mary is blond” and “Mary is tall and thin, and blond” amongst the truths connected by the grounding concept; otherwise, it can be seriously doubted that we understand each other when talking about grounding.

Let us now try to understand what kind of grounding relation the truths “Mary is thin and tall”, “Mary is blond” and “Mary is tall and thin, and blond” enjoy. As long as we think that “Mary is tall and thin” and “Mary is blond” are the strict, full and immediate grounds of “Mary is tall and thin, and blond”, we should think the same of “Mary is thin and tall” and “Mary is blond”: there is indeed no difference between these two pairs of truths that could possibly lead us to classify the former according to one type of grounding, and the latter according to another type.

Let us formalize what we have said in the following way:

- “Mary is tall and thin” and “Mary is blond” are formalized by $p \wedge q$ and r,
- “Mary is thin and tall” and “Mary is blond” are formalized by $q \wedge p$ and r,
- “Mary is tall and thin, and blond” is formalized by $(p \wedge q) \wedge r$.

According to the logics **FG** and **CG**, $p \wedge q$ and r are the full and immediate grounds of $(p \wedge q) \wedge r$: in the former case this is an instance of one of the axioms of Fig. 15.1, in the latter case this is due to the rule $\wedge I$. On the other hand, in neither of the logics **FG** and **CG**, are $q \wedge p$ and r the full and immediate grounds of $(p \wedge q) \wedge r$: indeed in the former case, there is no axiom the describes such a relation, in the latter case there is no rule that allows us to infer such a link.[8] Given what we have said in the previous paragraph, this is an important defect of the logics **FG** and **CG**: by means of these logics we should indeed be able to prove all correct grounding relations, so the lack of the one holding amongst $q \wedge p$ and r and $(p \wedge q) \wedge r$ counts as a serious flaw of these theories.

Let us note that the problem raised by the relation among the truths $q \wedge p$, r and $(p \wedge q) \wedge r$ for the logics **FG** and **CG** is just a small example of a more general situation. In the rest of the section we describe in detail this situation. Consider a truth A; A is said to be *a-c equiv* to another truth B if, and only if, B can be obtained from A by applications of associativity and commutativity of conjunction and disjunction.[9] For example, if A is of the form $E \wedge F$, then the formula $F \wedge E$ is *a-c equiv* to it. If A is of the form $\neg((B \vee C) \wedge (D \vee F))$ the formulas $\neg((C \vee B) \wedge (D \vee F))$, $\neg((B \vee C) \wedge (F \vee D))$, $\neg((C \vee B) \wedge (F \vee D))$ are *a-c equiv* to it. If A is of the

[8] Worse, if we are not mistaken, in the logics **FG** and **CG** $q \wedge p$ and r cannot be shown to be the grounds of $(p \wedge q) \wedge r$ *tout court*.

[9] We omit the formal definition of this notion for the sake of brevity.

form $((B \vee C) \vee (D \vee F))$, then the formulas $((B \vee D) \vee (C \vee F))$, $((D \vee B) \vee (F \vee C))$, $((B \vee F) \vee (D \vee C))$ are all *a-c equiv* to it. The central remark here is the following. If grounding is defined as an explanatory *objective* relation amongst truths, then the order and the way in which these truths are arranged should not condition the grounding relation in any way: neither in the very existence of the grounding relation nor in the type this grounding relation is. Therefore, not only the full and immediate grounds of A are the truths M indicated by the axioms of Fig. 15.1 of **FG** (or, equivalently, by the rules of Fig. 15.4 of the logic **CG**), but also, for any set of truths N and any truth C such that C is *a-c equiv* to A, if N is the full and immediate ground of C according to the axioms of Fig. 15.1 of **FG** (or, equivalently, according to the rules of Fig. 15.4 of the logic **CG**), then N is also the full and immediate ground of A.

Let us consider a few examples that support what we have just said. Consider the truths $\neg\neg((p \wedge q) \wedge t)$ and $(p \wedge q) \wedge t$: according to the logics **FG** and **CG** these truths stand in a full and immediate grounding relation. Consider now the following truths: $(q \wedge p) \wedge t$, $(p \wedge t) \wedge q$, $(t \wedge p) \wedge q$, $(q \wedge t) \wedge p$, ... It seems difficult to support the claim that in a grounding hierarchy $(p \wedge q) \wedge t$ and each of these latter truths occupy different positions. On the contrary, as long as $(p \wedge q) \wedge t$ is a (full and immediate) ground of the truth $\neg\neg((p \wedge q) \wedge t)$, each of these truths $(q \wedge p) \wedge t$, $(p \wedge t) \wedge q$, $(t \wedge p) \wedge q$, $(q \wedge t) \wedge p$, ... should be a full and immediate ground too.

Consider now the truth $((p \vee q) \vee t) \wedge r$; everybody agrees that $((p \vee q) \vee t)$ and r are the full and immediate grounds of this truth. But then everybody should also agree on the fact that $((q \vee p) \vee t)$ and r are the full and immediate grounds of $((p \vee q) \vee t) \wedge r$ too: between $((p \vee q) \vee t)$ and r on the one hand, and $((q \vee p) \vee t)$ and r on the other hand, there is no difference that would justify the claim that the former are (full and immediate) reasons of the truth $((p \vee q) \vee t) \wedge r$, and the latter are not. The same holds for other pairs of truths such as: $((q \vee t) \vee p)$ and r, $((t \vee q) \vee p)$ and r, $((t \vee p) \vee q)$ and r ...: we cannot at the same time claim that $((p \vee q) \vee t)$ and r are the full and immediate grounds of $((p \vee q) \vee t) \wedge r$, and each of these pairs is not.

Note that at a first glance one might be lead to think that the associativity of conjunction and disjunction is less natural than the commutativity of conjunction and disjunction in the classification of truths linked by the full and immediate grounding relation. Under such a perspective, some doubts might arise about what has been argued for in this section. In order to remove such doubts, let us add the following observations. Instead of considering the relation of grounding from the grounds perspective, let us consider it from the perspective of the conclusions. Consider indeed two truths like $(A \vee B) \vee C$ and $A \vee (B \vee C)$. These two truths are logically equivalent; plus they are made up by the same connectives used a same number of times. Finally, they are composed by the same propositional atoms. In a grounding hierarchy, which has a pure ontological nature, these two truths should be considered as identical. Indeed the only difference between them is the order in which the propositional atoms are grouped; such a difference is certainly significant from an epistemic point of view but basically irrelevant from a metaphysical one. From this latter point of view, there is no important difference between the two truths and therefore it seems hard to deny that they do not have the same full and immediate grounds.

We can thus conclude that the relation of full and immediate grounding is in a certain sense closed under associativity and commutativity of conjunction and disjunction and this characteristic is neglected by the logics **FG** and **CG**. Axioms and rules should be added to take into account these important grounding relations.

15.9 Conclusions

In this paper we have focussed on the two most recent logics of grounding: the one introduced by Fine in 2012, and the one introduced by Correia in 2014. Both these logics have the great merit and the undoubted virtue of contributing to the formal study of the principles concerning the grounding relation: our intuitions on grounding are systematized in a rigorous and precise treatment. In this paper we have tried to show in what way the formal principles put forward by Fine and Correia should be in some cases changed, in other cases enhanced. The intention has been to lay the basis for future debates and developments.

References

Avron, A. (1991). Simple consequence relation. *Information and Computation*, *92*, 105–139.

Batchelor, R. (2010). Grounds and consequences. *Grazer Philosophische Studien*, *80*, 65–77.

Betti, A. (2010). Explanation in metaphysics and Bolzano's theory of ground and consequence. *Logique et analyse*, *211*, 281–316.

Bolzano, B. (1973). *Theory of science: A selection, with an introduction*. Dordrecht: D. Riedel.

Bolzano, B. (1996). Contributions to a more well founded presentation of mathematics. In W. B. Ewald (Ed.), *From Kant to Hilbert: A source book in the foundations of mathematics* (pp. 176–224). Oxford: Oxford University Press.

Casari, E. (1987). Matematica e verità. *Rivista di Filosofia*, *78*(3), 329–350.

Correia, F. (2010). Grounding and truth-functions. *Logique et Analyse*, *53*(211), 251–279.

Correia, F. (2014). Logical grounds. *Review of Symbolic Logic*, *7*(1), 31–59.

Correia, F., & Schnieder, B. (2012). Grounding: An opinionated introduction. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding* (pp. 1–36). Cambridge: Cambridge University Press.

Daily, C. (2012). Scepticism about grounding. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding* (pp. 81–100). Cambridge: Cambridge University Press.

de Rosset, L. (2013). What is weak ground? *Essays in Philosophy*, *14*(1), 7–18.

Fine, K. (2012a). Guide to ground. In F. Correia & B. Schnieder (Eds.), *Metaphysical grounding* (pp. 37–80). Cambridge: Cambridge University Press.

Fine, K. (2012b). The pure logic of ground. *Review of Symbolic Logic*, *25*(1), 1–25.

Gentzen, G. (1935). Untersuchungen über das logische schließen. *Mathematische Zeitschrift*, *39*, 176–210.

Indrezejczak, A. (1997). Generalised sequent calculus for propositional modal logics. *Logica Trianguli*, *1*, 15–31.

Paoli, F. (1991). Bolzano e le dimostrazioni matematiche. *Rivista di Filosofia*, *LXXXIII*, 221–242.

Poggiolesi, F. (2008). A cut-free simple sequent calculus for modal logic S5. *Review of Symbolic Logic*, *1*, 3–15.

Poggiolesi, F. (2010). *Gentzen calculi for modal propositional logic*. Dordrecht: Springer.

Poggiolesi, F. (2012). On the importance of being analytic. The paradigmatic case of the logic of proofs. *Logique et Analyse*, *219*, 443–461.

Poggiolesi, F. (2013). From a single agent to multi-agent via hypersequents. *Logica Universalis*, *3*, 443–461.

Prawitz, D. (1965). *Natural deduction. A proof theoretic study*. Stockholm: Almquist and Wiskell.

Rumberg, A. (2013). Bolzano's theory of grounding against the background of normal proofs. *Review of Symbolic Logic*, *6*(3), 424–459.

Schaffer, J. (2009). On what grounds what. In D. Manley, D. J. Chalmers, & R. Wasserman (Eds.), *Metametaphysics: New essays on the foundations of ontology* (pp. 347–383). Oxford: Oxford University Press.

Schnieder, B. (2011). A logic for 'Because'. *The Review of Symbolic Logic*, *4*(03), 445–465.

Sebestik, J. (1992). *Logique et mathematique chez Bernard Bolzano*. Paris: J. Vrin.

Tatzel, A. (2001). Proving and grounding: Bolzano's theory of grounding and Gentzen's normal proofs. *History and Philosophy of Logic*, *1*, 1–28.

Tatzel, A. (2002). Bolzano's theory of ground and consequence. *Notre Dame Journal of Formal Logic*, *43*(1), 1–25.

Troelstra, A. S., & Schwichtenberg, H. (1996). *Basic proof theory*. Cambridge: Cambridge University Press.

Chapter 16
Computability, Finiteness and the Standard Model of Arithmetic

Massimiliano Carrara, Matteo Plebani and Enrico Martino

Abstract This paper investigates the question of how we manage to single out the natural number structure as the intended interpretation of our arithmetical language. Horsten (2012) submits that the reference of our arithmetical vocabulary is determined by our knowledge of some principles of arithmetic on the one hand, and by our computational abilities on the other. We argue against such a view and we submit an alternative answer. We single out the structure of natural numbers through our intuition of the *absolute* notion of finiteness.

Keywords Computational structuralism · Finiteness · Standard model of arithmetic

16.1 Introduction

Claiming that arithmetic is about the natural number structure (N) may sound trite. But since every first-order version of arithmetic has models that are not isomorphic to each other, it is natural to wonder how we manage to single out a class of isomorphic models, i.e. the natural number structure, as the intended interpretation of our arithmetical language.

Matteo Plebani acknowledges financial support of his work by the Spanish Ministry of Economy and Competitivity and FEDER through the project: *The explanatory function of abstract objects: their nature and cognoscibility*, FFI2013-41415-P.

M. Carrara (✉) · E. Martino
University of Padua, Padua, Italy
e-mail: massimiliano.carrara@unipd.it

E. Martino
e-mail: enrico.martino@unipd.it

M. Plebani
University of Santiago de Compostela, Santiago de Compostela, Spain
e-mail: plebani.matteo@gmail.com

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_16

According to computational structuralism, the position defended by Horsten (2012) (see also Halbach and Horsten 2005), there are two aspects of our arithmetical practice that contribute to fix the natural number structure as its intended interpretation (Horsten 2012, p. 280):

(I) A *know-that component*: our endorsement of some arithmetical principles, as expressed for instance in Peano Arithmetic (PA).

(II) A *know-how component*: our use of an algorithm of computation to calculate sums.

The major novelty of the proposal lies in step (II), whose point, however, is very easy to be misunderstood. In particular, it might not be always clear exactly which question computational structuralists are addressing. A careful scrutiny of their proposal is thus called for.

16.2 The Question

The question **Q** we are going to address has been nicely phrased by Horsten (2012, p. 276):

(Q) *How do we* manage *to single out the structure of the natural numbers as the* interpretation *of our arithmetical vocabulary?*

Horsten pinpoints that **Q** is not to be understood as the *skeptical* question whether we succeed in singling out the natural number structure as the target of our arithmetical practice. **Q** *presupposes* that we do succeed in that, and just asks *how* this is possible. In Horsten's words:

> We could, with Skolem or Putnam [...] deny that arithmetic has an intended interpretation that is unique up to isomorphism [Putnam 1980]. But I will do the opposite. I will presume that we can isolate the natural-number structure in our referential practice. So I will have no qualms about referring to the natural numbers and operations on them in my philosophical account. The question addressed in this article is how we have managed to refer to the natural-number structure: *how has our reference to the natural-number structure come about?* (Horsten 2012, p. 277)

We think, however, that the relevant question is: what is the standard model of PA? What characterizes it among the infinite plurality of models? Once the peculiar feature of such a model has been characterized, no further question arises about how to refer to it: we refer to it as to the model so characterized. On the other hand, the mere presupposition that we are able to single out a certain privileged model is obviously of no help for identifying it.

16.3 The Computational Structuralist Answer

In a nutshell, Horsten's answer to **Q** is:

"(Thesis 2) The reference of our arithmetical vocabulary is determined by:

(I) our principles of arithmetic together with
(II) our use of an algorithm of computation." (Horsten 2012, p. 280, numeration added)

From (I) and (II) we can reach the conclusion that N is the intended interpretation of our arithmetical vocabulary in the following way: First, our endorsement of some principles of arithmetic, as codified, say, in the axioms of Peano Arithmetic, singles out a class of models (those satisfying the axioms) as possible interpretations of our mathematical practice (element I). This is not enough to rule out non-standard models, of course. But we notice that there is a further aspect in our arithmetical practice: we *use* an algorithm to compute sums (element II). Also this aspect of our arithmetical practice should be reflected in the interpretation of our arithmetical vocabulary: the function associated to '$+$' in the model must be computable.

By the **Church-Turing Thesis**, if a function is intuitively computable, it is Turing-computable. And by **Tennenbaum's theorem,** all models of PA, in which the addition function is Turing-computable, are isomorphic to N (Tennenbaum 1959). So it follows by the **Church-Turing Thesis** and **Tennenbaum's theorem** that all models of PA, in which the addition function is computable, are isomorphic to N (see Horsten 2012, p. 280).

Notice that Horsten makes clear that the notion of computability relevant here is a pre-theoretical one. It is the kind of intuitive notion that we try to model mathematically using notions like Turing computability, but it is not itself a theoretical notion.[1] It is important to stress this point in order to understand Horsten's own position. One version of the so-called 'just more theory' objection (Putnam 1980) applies to the suggestion that the defining feature of the standard model is that in such model addition is a computable function. The objection is that, adding to the axioms of PA the translation into first-order language of the claim that addition is recursive, one still obtains a theory that has non-standard models. One thing Horsten says in reply to this objection (Horsten 2012, p. 287) is that all that the existence of non-standard models shows in this case is that the intuitive notion of computability cannot be completely formalized; but this does not make it an illegitimate notion: the fact that invisible gases cannot be seen is no reason to believe there are no invisible gases, to use a nice example from Liggins (2012). Moreover, Horsten's appeal to an intuitive, pre-theoretical notion is also important for the dialectic of this paper, because we

[1] "The sense in which addition is algorithmically computable is not to be understood as Turing-computability or μ-recursiveness. What is operative here is the pre-theoretical practical sense of computability, which has motivated the mathematical definitions of computability. It is a practical notion of computing on strings of symbols" (Horsten 2012, p. 278).

will also appeal to an intuitive, pre-theoretical notion, more fundamental than that considered by Horsten.[2]

The two main ingredients of Horsten's answer are given by a formal theory like PA and our ability to compute sums, in the pre-theoretical sense of *computation*. In the next section we are going to argue that both these aspects rest on something more fundamental: our ability to generate the relevant syntactical entities that constitute a formal theory like PA and are the basis on which the addition algorithm works. This, in turn, rests on our ability to grasp a primitive notion of finiteness. Observe that what is relevant in our perspective is that *absolute finiteness* is presupposed for the understanding of a formal system as well of a Turing machine.

16.4 Problems with Computational Structuralism

Let us start with (I): our endorsement of some principles of arithmetic.

The knowledge of the principles of arithmetic presupposes, in turn, the knowledge of the *formal* arithmetical language L (including terms and formulas). Horsten says explicitly that the language L at issue is that of first-order Peano arithmetic PA. The problem arises: how can one understand the inductive definition of the arithmetical language?

It is often remarked that syntactical notions like numerals, formulas and so on are informally defined in terms of the notion of *finiteness* (see Field 2001, p. 338): a numeral is defined as a *finite* sequence of signs, formulas are *finite* sequences of symbols meeting certain conditions. And the idea that all these entities can be inductively defined relies on the idea that the inductive clauses describe a procedure for generating each of them in a *finite* number of steps. So the endorsement of the axioms of PA presupposes the notion of *finiteness.*

Something similar holds for (II), i.e. our use of an algorithm for computing sums. First of all, an algorithm is computable in a *finite* sequence of steps. Moreover, computation, as Horsten says, is something we perform on symbols (2012, p. 278). Our use of an algorithm for computation is possible only on the condition of understanding which objects this algorithm works on. In the case of the sum algorithm, one takes numerals as inputs and outputs. And since the sum must operate on every pair of numerals, the notion of sum presupposes the *general* notion of numeral and hence the general notion of a *finite string of signs*. This is more fundamental than the items (I) and (II).

Now, the notion of finiteness at issue is just what serves the purpose of singling out the standard model: this is characterized by the fact that every element has finitely many predecessors. The set of numerals, structured with the usual arithmetical operations, is a paradigmatic example of the standard model of arithmetic.

[2]To show that the notions of intuitive computability/finiteness are legitimate it is not enough to show that they are not illegitimate: one needs also positive reasons to accept them. We provide such reasons (for the finiteness case) in Sects. 16.5 and 16.6 (Thanks to an anonymous referee for pressing us on that).

Question **Q** can, therefore, reformulated as follows:

Q*: how can we grasp the notion of finiteness?

We will discuss this question in the next section.

16.5 The Absolute Notion of Finiteness

As it is well known, the notion of finiteness defined in set theory is relative to a model of set theory. In each model of set theory one defines the set of natural numbers and therefore a notion of finiteness. But the sets of natural numbers of two different models may be non-isomorphic. Of course, one could define the standard model of arithmetic as the structure of the natural numbers of a *standard model of set theory*, which should catch our pre-theoretical notion of finiteness. But that would lead us to the much more problematic question: what is the standard model of set theory? Similarly for the notion of finiteness defined in second order logic (see Weston 1976).

The notion of finiteness we are concerned with is a *primitive absolute* (i. e. independent of any set-theorethical model) notion that no axiomatic system is adequate to capture. It is just this inadequacy that may suggest the skeptical doubt that the alleged notion of *absolute* finiteness is illusory. We want to argue that such a doubt is untenable.

The crucial point is that one can recognize the existence of non-standard models of arithmetic only after having introduced the formal language of arithmetic and having grasped a structure where the axioms of arithmetic hold.

Now, as already stressed, the comprehension of the formal language rests on the intuition of a *finite string* of signs. This intuition is primitive and transcends any practical mastery of constructing particular strings. Such strings, which C. Parsons calls “quasi-concrete” (Parsons 1990), as types of spatial objects, are immediately given to intuition. They are, according to Hilbert, extra-logical objects presupposed by logic, objects whose intelligibility is an essential condition for human reasoning:

> As a condition for the use of logical inferences 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. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that neither can be reduced to anything else nor requires reduction. This is the basic philosophical position that I consider requisite for mathematics and, in general, for all scientific thinking, understanding and communication. And in mathematics, in particular, what we consider is the concrete signs themselves, whose shape, according to the conception we have adopted, is immediately clear and recognizable. (Hilbert 1926, 376).

It is clear in this passage that Hilbert’s finitism is grounded on the notion of absolute finiteness. Ironically, Hilbert’s finitism is often considered as formalizable in primitive recursive arithmetic, in spite of the fact that no axiomatic system can catch the intended notion of finiteness.

In virtue of this notion, we are able to regard the usual inductive definition of the formal system of arithmetic as a description of a process generating all the intended syntactical entities. Our grasp of such an ideal process, which leads to any string in a *finite* number of steps, rests on the intuition of *a finite number of steps* and of *a finite sequence of signs*.

Consider, for example, the inductive definition of a string of strokes:

(1) One stroke is a string of strokes;
(2) adding a stroke to a string of strokes one obtains a string of strokes.

To understand the above definition as a description of a generating process, one must possess a priori the intuitive notion of a finite string of strokes, presupposed by clause 2. Otherwise one has to understand the above clauses as mere axioms. In this case what the inductive definition says is only that strings of strokes are certain things satisfying the above clauses. Then the clauses are capable of non-standard interpretations and the notion of string of strokes is indeterminate.

As to the axioms of PA, the evidence for the induction schema rests in turn on the intuitive notion of finiteness. Lacking that, nothing would guarantee that any property, which holds for 0 and is inherited from any number to its successor, holds for all numbers. The intuition that any number is reachable, starting from 0, in a finite number of steps, is essential for the evidence that any number inherits the property at issue from 0. For this reason the evidence of induction rests on the notion of absolute finiteness.

Observe that absolute finiteness plays this role even when the principle of induction is expressed within an informal language. Concerning this point, it should be noted that, though the axioms of arithmetic fail to determine the standard model, their evidence presupposes the grasp of the standard model.[3]

Our claim is that the standard model is characterized by the fact that every natural number has finitely many predecessors, where the notion of finiteness here involved is absolute and primitive. This means that it cannot be *defined* in terms of more elementary notions.

Perhaps Horsten could object that the notion of finiteness can be grasped, in turn, through a training in practical computation. We do not deny the possibility that the intuition of a finite string of signs is acquired by an act of abstraction suggested by the computational practice. In this case, however, such an abstraction is already suggested by the mere practice of counting; the practice of adding, invoked by Horsten, is overwhelming.[4]

[3]Observe that a skeptical, who denies the existence of the standard model, should deny the evidence of the principle of induction.

[4]It is still worth noticing one point. Horsten thinks that every attempt to characterize the standard model by making reference to our ability to count up to every natural number works only if certain empirical conditions are met. Teaching how to count and adding that all the natural numbers can be generated by counting succeeds in excluding non-standard models only if time has the right structure (Horsten 2012, p. 283). Indeed, if time had an irregular structure, the procedure to generate natural numbers by counting would give unintended results. But the same holds for the addition algorithm,

16.6 Skepticism Reconsidered

Button and Smith (2012) argue that neither *computability* nor *finiteness* can answer skeptical challenges about our ability to isolate the standard model of arithmetic. As there are non-standard interpretations of our arithmetical vocabulary, they argue, there are non-standard interpretations of the theories formalizing our notion of computability and finiteness.

If the skeptic takes for granted that we have a well-determined arithmetical language and denies that this has a privileged interpretation, then we can reject her position since, as we saw, the standard model is exemplified by the numerals.

On the other hand, a more radical skeptic maintains that the inductive definition of formulas and terms is to be regarded not as a description of a generating process but—in turn—as an axiomatic system that fails to single out the intended structure of the syntactical entities. This position has been illustrated above with the example of the definition of a finite string of strokes.

This is a radical form of skepticism, namely skepticism about the possibility of singling out a well-determined infinite formal language. Such a position destroys the very aim of formalization to make precise the fundamental mathematical notions of formula and proof (see Field 2001, p. 338).

16.7 Conclusions

In this paper we have considered the question of how can we single out the standard model of arithmetic. We have analyzed Horsten's answer based on (1) our knowledge of some principles of arithmetic and (2) our mastery of the algorithm for computing sums. We argued that the understanding of item (1) rests on the absolute primitive notion of finiteness not formalizable by any axiomatic system. This notion is all that is required to single out the standard model of arithmetic. Besides it is essential even for understanding any infinite formal system. As to item (2), absolute finiteness transcends any computational practice. And if it is grasped through an act of extrapolation on our mastery of manipulating certain strings of signs, then it is already grasped by an act of abstraction on the mere mastering of counting.

(Footnote 4 continued)
mastery of which is, according to Horsten, one of the factors that determine the reference of our arithmetical vocabulary. For the addition algorithm to work, nature must play its part by having a space dimension structured in the right way: if space contracted when concatenating two numerals, then the result of summing m and n would not be the intended one. Anyway, we agree that the intuition of the generating process of the numerals involves a naive intuitive notion of time. For our purposes, it doesn't matter whether our naive notion of time is in agreement with the real structure of time or not.

References

Button, T., & Smith, P. (2012). The philosophical significance of Tennenbaum's theorem. *Philosophia Mathematica*, *20*(1), 114–121.

Dean, W. (2013). Models and computability. *Philosophia Mathematica*, *22*(2), 143–166.

Field, H. (2001). *Truth and the absence of fact*. Oxford: Oxford University Press.

Halbach, V., & Horsten, L. (2005). Computational structuralism. *Philosophia Mathematica*, *13*(3), 174–186.

Hilbert, D. (1926). Über das Unendliche. *Mathematische Annalen, 95*, 161–190. Lecture given Münster, June 4, 1925. English translation in van Heijenoort (1967, pp. 367–392).

Horsten, L. (2012). Vom Zahlen zu den Zahlen: On the relation between computation and arithmetical structuralism. *Philosophia Mathematica*, *20*(3), 275–288.

Liggins, D. (2012). Weaseling and the content of science. *Mind*, *121*(484), 997–1005.

Parsons, C. (1990). The structuralist view of mathematical objects. *Synthese*, *84*(3), 303–346.

Putnam, H. (1980). Models and reality. *Journal of Symbolic Logic*, *45*, 464–482.

Tennenbaum, S. (1959). Non-archimedean models for arithmetic. *Notices of the American Mathematical Society*, *6*(270), 44.

van Heijenoort, J. (Ed.). (1967). *From Frege to Gödel. A source book in mathematical logic*, 1897–1931, Cambridge, MA: Harvard University Press.

Weston, T. (1976). Kreisel, the continuum hypothesis and second order set theory. *Journal of Philosophical Logic*, *5*(2), 281–298.

Chapter 17
The Significance of a Categoricity Theorem for Formal Theories and Informal Beliefs

Samantha Pollock

Abstract This paper considers the existing literature on what a categoricity theorem *could* achieve, and proposes that more clarity and explicit admission of background beliefs is required. The first claim of the paper is that formal properties such as categoricity can have no philosophical significance whatsoever when considered apart from informal, philosophical beliefs. The second is that we can distinguish two distinct types of philosophical significance for categoricity. The paper then highlights two consequences of this analysis. The first is a potential source of circularity in Shapiro's wider project, that arises out of what appear to be arguments for both kinds of philosophical significance with respect to categoricity. Rather than a knock-down objection to Shapiro's philosophy of mathematics, the discussion of this potential circularity is intended to demonstrate that implicit claims surrounding the significance of categoricity can lead to philosophical missteps without due caution. The second outcome is that, as an initial case study, categoricity has limited significance for the semantic realist.

Keywords Categoricity · Formalisation · Realism

17.1 Introduction

Since Dedekind's (1888) categoricity theorems for arithmetic and real analysis, the question of what the philosophical significance of a categoricity theorem might be has tended only to be answered with respect to specific mathematical theories.[1] The aim of this paper is to present a framework for better examining the philosophical significance of categoricity in relation to *particular* mathematical theories, by *generalising* the backdrop to such investigations. The first half of the paper will focus on clarifying the distinction between 'formal' and 'informal' in mathematics and

[1] There are notable exceptions. See Corcoran (1980), Meadows (2013) and Button and Walsh (2015).

S. Pollock (✉)
University of Bristol, Bristol, UK
e-mail: sp8777@bristol.ac.uk

F. Boccuni and A. Sereni (eds.), *Objectivity, Realism, and Proof*,
Boston Studies in the Philosophy and History of Science 318,
DOI 10.1007/978-3-319-31644-4_17

arguing that, given this distinction, the philosophical significance of categoricity has all to do with the relation between our beliefs about informal theories (e.g. 'arithmetic') on the one hand, and the properties of corresponding formal theories (e.g. PA^2) on the other. After distinguishing two different potential sources of significance for a categoricity theorem, the second half of the paper explores some consequences of this distinction.

An important preliminary is the definition of 'categoricity'. The property is defined in model-theoretic terms as follows:

Categoricity

A mathematical theory is *categorical* iff for all models $\mathcal{M}_i, \mathcal{M}_j$ that satisfy the theory, $\mathcal{M}_i \cong \mathcal{M}_j$.

It would be a category mistake to ask whether categoricity holds of arithmetic per se; that question only applies to formal, axiomatic theories of arithmetic with model-theoretic semantics. A related question can be asked of arithmetic, though, namely: to what extent are the natural numbers unique? It is crucial to recognise that these two questions are distinct. One is a formal question for the model-theorist, whilst the other is an informal question for the philosopher and has nothing *necessarily* to do with models. There is nothing to a priori rule out a fact of the matter about the uniqueness of the natural numbers that is independent of model theory. The following section elaborates on this distinction.

17.2 Formal and Informal Mathematical Theories

The properties of formal theories cannot be of much philosophical interest when viewed apart from any informal input. That is to say, 'It should not be supposed that the formalism can grind out philosophical results… There is no mathematical substitute for philosophy' (Kripke 1976, p. 414). So, if the formal property of categoricity is to have any philosophical significance, we must consider both the formal mathematical theories of which it may hold, and the informal mathematical reasoning about which we may have philosophical beliefs. The following distinction between formal and informal mathematical theories relates to work in the philosophy of mathematics on informal rigour and proof, axiomatisation versus formalisation, and the epistemological and methodological claims that arise from it. However, it is not itself intended as a contribution to epistemology or methodology, but rather as a 'lay of the land' in order to cleanly and explicitly mark out the potential place for categoricity to play its oft-cited role in matching up the formal with the informal in mathematics.

17.2.1 Informal Mathematical Theories

The practice of mathematics pre-dates[2] the axiomatic method; mathematicians speak and reason in informal terms; and the schooled layperson is acquainted with arithmetic even if she has never seen a formal presentation of the Peano axioms, or considered which deductive system and semantics are appropriate for its formalisation. What we do pre-axiomatisation, in every day practice, and in the classroom is informal mathematical reasoning. In Lakatos' terms, which take careful consideration of the historical development of mathematics, there are not just formal proofs, but informal (either 'pre-formal' or 'post-formal') proofs, too (Lakatos 1978).

To get from this observation to the claim that there are such things as informal mathematical theories, all we have to propose is that the observed informal mathematical reasoning going on around us is minimally uniform, coherent, and divisible into distinct, identifiable areas that we can think of as 'theories'. A seemingly natural identity criterion for informal mathematical theories would invoke subject matter. This is because, at first glance, it looks as though informal mathematical reasoning can be categorised according to distinct mathematical subject matters—real numbers, groups, sets, fields, and so on. As for uniformity and coherence, one thing that all of this informal mathematical reasoning has in common is that it does not contain non-mathematical, or mixed, propositions. These are very minimal requirements on what constitutes an informal mathematical theory so as not to over-formalise the very idea. From these simple observations, then, we can suggest the following definition:

Informal Mathematical Theory

An *informal mathematical theory* is a collection of purely mathematical assertions in natural language, with a common subject matter.

The natural language assertions of an informal theory must be 'purely mathematical' in the sense that 'there is no greatest prime number' is purely mathematical, but 'I prefer 17 to 9' is not. We can accept without too much worry that: (a) informal mathematical assertions either share, or do not share, the same subject matter; and (b) we can, in general, distinguish between these two cases for any given pair of assertions. This is so, regardless of what subject matters are, and what their ontological status is. There just *is* a difference between informal arithmetical assertions and informal geometrical assertions: they assert $x, y, z, \ldots$ of different subject matters.

One might object for a couple of reasons. Firstly, one could argue that arithmetic and set theory are different mathematical theories, but that the subject matter of arithmetic is nevertheless sets, on a view of set theory as a foundation for mathematics. This can be thought of as a matter of perspective, however. From the philosopher's metamathematical vantage point, we can make foundational claims like this, but from the mathematician's theory-level perspective, the subject matter of arithmetic is the natural numbers whilst the subject matter of set theory is sets. On the other

[2]This is not intended strictly; certain of what many regard as axiomatisations predate a large body of mathematical practice, such as Euclid's geometry. Still, we can suppose that Euclid informally reasoned about space before fixing his postulates.

hand, one might object that there is overlap in subject matter between, for example, arithmetic and real analysis since the natural numbers are included among the real numbers. However, overlap does not constitute identity; the real numbers, though they *include* the natural numbers, are not the subject matter of arithmetic. The same can be said of other seeming overlaps of subject matter. For instance that the real numbers include the rationals, which include the integers, which include the natural numbers, just as we might say the subject matter of biology includes the subject matter of chemistry, which includes the subject matter of physics—whilst maintaining that these are different scientific theories with identifiably distinct subject matters.

Mathematical theories and areas of research are, of course, not islands, just as these and other mathematical subject matters are not remote from one another. Yet, inasmuch as the work done in a mathematics department, or indeed the chapters of a student's mathematics syllabus, can be delineated according to theory, we may accept that those theories can be distinguished from one another on the basis of what they are about. This is decided before axiomatisation or formalisation take place, for axioms seek to determinately and completely describe a subject matter, and formalisation involves specifying the domains of models in which those axioms hold or fail to hold. In Antonutti Marfori's words, theories are axiomatised when their 'fundamental or primitive notions are singled out and the other notions are defined from them, and its fundamental or primitive propositions are identified as axioms and the other propositions are derived from them by chains of deductions'; later, axiomatised theories are formalised when truth can be preserved through inference, by their being 'expressed by a purely syntactical language' (2010, pp. 261–62). She notes, too, that theories can be axiomatised either formally or informally. We might expect, therefore, that the typical development of a mathematical theory would progress from: (1) an informal theory as defined above; to (2) an informal axiomatisation that separates the primitive from the non-primitive assertions; perhaps to (3) a formal axiomatisation in which the primitive propositions are translated into a formal language; to, finally, (4) a ('full') formalisation as described below (Sect. 17.2.3), which adds a choice of deductive system and semantics to the set of axioms in a formal language.

17.2.1.1 Philosophical Beliefs About Informal Theories

Several philosophers have attributed the properties of being able to 'pick out' particular mathematical objects or concepts, of being 'about' a unique model or structure, and of having an 'intended interpretation', to certain mathematical theories. Moreover, they often attach importance to these properties. The following is just a small sample of where this kind of sentiment appears in the literature:

> Formalized mathematics may… serve as a mediator of sorts between different foundational views. But for this very reason it does not fully capture the view that underlies ordinary mathematical practice – in as much as the practice implies a particular structure that constitutes *the* subject matter of the inquiry, "what it is all about".
> (Gaifman 2004, p. 3)

> …we know much better what we are talking about when we are doing arithmetic than when we are doing set theory. At least we are more disposed to admit that we cannot so easily single out the intended models of set theory than to admit that arithmetic is also about nonstandard models.
> (Halbach and Horsten 2005, p. 185)

> Informal rigour is how we know what we are talking about. I also want to argue that the categoricity results for e.g. theory of the natural numbers and theory of sets show that these axiomatisations capture *fully* the intended notion.
> (Isaacson 2011, p. 26)

> Second-order formulations of set-theory do not capture *the* intended interpretation. This is worrying.
> (Melia 1995, p. 129)

It is precisely the kind of properties mentioned by these philosophers that are invoked in discussions of the significance of categoricity. As such, these properties deserve a little more analysis. The first thing to say is that sometimes we believe that informal theories are about a *unique* subject matter. The debate over the ontological status of natural numbers, sets, and so on is another question. We can clearly single out two very basic beliefs about a given informal theory T_I, to which we may or may not subscribe, and which obviously underlie the notions of aboutness, intendedness and 'picking out', etc.:

Uniqueness

The subject matter of T_I is unique (e.g. the subject matter of analysis is *the* structure of the real numbers).

Existence

The subject matter of T_I objectively exists (e.g. the real numbers exist as a Platonic structure).

Seemingly, all of the notions alluded to in the above quotations involve a commitment to some form of the Uniqueness thesis. We think that some theories pick out unique structures, are about unique subject matters, have uniquely intended interpretations, whilst others do/are/have not. For example, we generally do not subscribe to any form of the Uniqueness thesis for group theory, yet many of us posit at least one variation of it with respect to arithmetic and real analysis. Some attribute it, at least linguistically, to set theory.[3] The Existence thesis, on the other hand, may or may not be wrapped up in our ascription of uniqueness to an informal mathematical theory. This is deeply related to what our background philosophical beliefs are: Do we think that something has to exist in order to be picked out? Do we think that something has to exist for it to be that which a theory is about? Must an interpretation exist in some ontologically robust way in order to be intended?

Familiar realism debates in philosophy of mathematics revolve around the extent, combination and qualification of philosophers' commitments to the Uniqueness and

[3]The latter claim is, of course, very controversial, and part of the controversy issues from the fact that there is no categoricity theorem for formalisations of set theory. Zermelo produced only a 'quasi-'categoricity result for ZFC. On the other hand, however, a handful of philosophers have offered informal arguments that they take to secure the uniqueness of at least the *concept* of set (see McGee 1997; Martin 2001; Welch 2012).

Existence theses. Consider three main forms of realism attributed by philosophers to informal mathematical theories: ontological, semantic and conceptual.

- The *ontological* realist about T_I believes that the subject matter of T_I exists objectively.
- The *semantic* realist about T_I believes that all mathematical propositions about the subject matter of T_I have a determinate truth value.
- The *conceptual* realist about T_I believes that all T_I-concepts are objective (or, at least, intersubjective).

Ontological realism obviously involves positing Existence in some sense, but there are more or less robust versions, and different brands take different stances on Uniqueness. For example, the *ante rem* structuralist is an ontological realist who believes that the natural numbers exist objectively as places in a unique Platonic structure. Meanwhile, the *in rebus* structuralist is also an ontological realist, yet believes that there is no unique Platonic structure over and above the various objectively existing "systems" that instantiate the would-be structure.

It is easy to see that semantic realism posits a form of Uniqueness in the sense of unique determination of truth value, though it does not posit any form of the Existence claim. However, if the semantic realist feels a pull from the Existence claim as a way of explaining *why* semantic realism about T_I is true (e.g. what confers truth values is the existence and nature of the objects of the propositions in question), then they may subscribe to another form of realism that asserts Existence. This is not necessary, however; a fictionalist about T_I *could* nonetheless be a semantic realist about T_I, simply by being a semantic realist about statements in a fiction, in the same way that one might believe every well-formed statement about Harry Potter is determinately true or false, even though Harry Potter does not exist outside of the author's fiction. It is easy to see how this belief might be unpopular, both in the case of the fictional wizard and in the case of the fictionalist's account of mathematics.

Perhaps a more plausible example of how a semantic realist about T_I might not assert ontological realism about T_I is nominalism about the ontology of T_I's subject matter, under a modal structuralist interpretation.[4] According to Hellman (1989), although mathematical objects like the number 2 do not exist and thus terms like '2' cannot refer to them, the modal-structuralist still seeks to maintain the classical 'truth-determinateness' of mathematical theories like arithmetic, via higher-order resources and considering the notion of '*possible* T_I-structures'. If one has good reason to be a semantic realist about arithmetic and is happy to accept a higher-order formalisation of the theory, but is not comfortable positing about 'the structure of the natural numbers', then one can maintain semantic realism without ontological realism with Hellman's modal-structuralist translations of the sentences of PA^2. Notice, of course, that categoricity appears to have just as important a role to play in Hellman's structuralism as it does in Shapiro's Platonist *ante rem* version, for both interpretations of structuralism treat categoricity as important for securing at least one reading of the Uniqueness thesis.

[4] I am grateful to an anonymous reviewer for suggesting this natural example.

According to the conceptual realist, meanwhile, mathematical theories are at least in part about concepts, which are objective. For Gödel (1951), these concepts have a rather robust ontological status, existing independently of our mathematical reasoning. His version of conceptual realism is a form of Platonism, in which talk of the concept of a mathematical object replaces talk of objects themselves; it is Platonism about the concept of set, rather than about sets themselves. For the Gödelian realist, we can have direct epistemic access to these concepts via what he calls 'mathematical intuition', a kind of perception. But one might hold a weaker form of conceptual realism about T_I, according to which all T_I-concepts are human constructs that are nevertheless objective in the weaker sense of being universally agreed upon. In sum, conceptual realism about a mathematical theory posits Uniqueness via this emphasis on objectivity or inter-subjectivity (e.g. Welch 2012; Martin 2012), and the Gödelian version builds Existence into the position.

In theory, we could combine any of these three types of realism, and their corresponding anti-realist theses, into a set of philosophical beliefs with respect to a given informal mathematical theory T_I; none rules out any other view besides its own opposite (e.g. ontological realism rules out ontological *anti*-realism, but no other thesis considered above). For example, one might follow Gödel and be both a conceptual and an ontological realist about T_I. One might be an ontological realist about T_I in virtue of first being a semantic realist about T_I, and believing that truth is determined in part by mathematical reality. Likewise, semantic realism might lead one to conceptual realism. On the other hand, as we saw, one might be a semantic realist but an ontological anti-realist. Thus, to hold any of these positions or their converses with respect to T_I is to make a decision on the extent to which one subscribes to Existence and Uniqueness for T_I, and to hold a combination of them and their converses is to make a broader set of such decisions. It is clear, as a result, that the Uniqueness thesis in particular is highly important to questions of realism and anti-realism in the philosophy of mathematics. It follows that categoricity, if it turns out to be significant with respect to Uniqueness, is a philosophically important property. Categoricity's potential for philosophical significance, therefore, is deeply tied to the informal notion of the uniqueness of mathematical subject matter, under e.g. an ontological, semantic or conceptual interpretation.

17.2.2 Formal Mathematical Theories

The textbook definition of a formal mathematical theory, call it T_F, consists of a set S of sentences in a formal language: the axioms of T_F. However, as Hintikka (2011, p. 72) says: 'there need not be anything wrong about this account of axiomatisation, but it is not the whole story.' Since it is pertinent to our discussion, we will also take it that specifying a formal mathematical theory involves specifying a particular deductive system, according to which further sentences can be derived from S. This reflects mathematical practice, since there would *be* no practice without deduction from axioms. All we stipulate here is that the allowable deductions with respect to T_F must be specified via an explicit choice of deductive system. There seems to be nothing very controversial in this stipulation.

Slightly more controversially, we take it that specifying a formal mathematical theory also involves choosing a semantics: a class of models that satisfy S. We are concerned specifically with categoricity and, as mentioned above, this notion is defined in terms of models. Thus, whether or not categoricity holds of a given formal mathematical theory depends on *which* class of models of that theory that we consider. Crucially for the present debate, a choice of semantics is not made arbitrarily; it reflects the interpretation that we assign to the quantifiers, operators, logical and non-logical symbols of a formal theory. The existence of a unique model (at least up to isomorphism) satisfying S is necessarily relative to such an interpretation. It is well known, for instance, that 'full' semantics for second-order logic interprets the second-order quantifiers as ranging over every subset of the powerset of the first-order domain, whilst a restricted version of Henkin semantics interprets them as ranging over strictly less than every subset of the powerset of the first-order domain.[5]

Therefore, ours is the following definition:

Formal Mathematical Theory

A *formal mathematical theory* T_F consists of a set S of sentences in a formal language $\mathcal{L}_{T_F}$, along with a specified deductive system and semantics.

In other words, a formal mathematical theory is for us fully specified. It tells us everything we need to know about syntactic and semantic consequence with respect to some mathematical domain.

17.2.3 Formalisations

Not every formal theory is intended to capture an informal theory; there is an intentional aspect to what we call a formalisation. As Wang (1955) notes, we can have several intentions when formalising an informal theory: systematising a field of mathematical study; improving our ability to communicate a field of mathematical study; clarifying and consolidating concepts, arguments and proofs; making mathematical practice rigorous; and approximating intuition. Considering these supposed purposes of formalisation, we can see that whilst every formalisation of an informal theory is a formal theory, not every formal theory is necessarily a formalisation of an informal theory. In Lakatos' opinion:

[5] The question of which of these semantics is 'standard' is itself a matter of debate. Shapiro (1985) takes full semantics to be standard for second-order logic, whilst Ferreirós (2011) takes restricted Henkin semantics to be 'just as natural as the preferred [full] one, or even more natural' (p. 383). One could think of full semantics as just a special case of Henkin semantics, in which we consider *every* subset of the powerset of the first-order domain.

> …we should speak of formal systems only if they are formalizations of established informal mathematical theories. No further criteria are needed. There is indeed no respectable formal theory which does not have in some way or another a respectable informal ancestor. (Lakatos 1978, p. 62)

Lakatos' prescriptivism is stronger than what is proposed here, but underlying the sentiment is the idea that formal theories and formalisations of informal theories are not necessarily the same thing, and should be distinguished. Another way to see the distinction between a formal theory and a formalisation is to consider properties that look like necessary properties of a formalisation, but not of formal theories. We propose that the following are necessary and jointly sufficient conditions for a formal theory T_F to be considered a formalisation of some informal theory T_I[6]:

- Adequate expressiveness
- Natural language tractability
- Truth preservation

Adequate expressiveness is the property that every well-formed, mathematical, natural language proposition p about the mathematical subject matter of T_I has a formal analogue ϕ in $\mathcal{L}_{T_F}$. In other words, the language of T_F must be 'complete' with respect to the language of T_I, such that it is at least as expressive as natural language (with specialist T_I-vocabulary) when it comes to talking about the mathematical subject matter in question. We require that there is nothing *purely mathematical* necessarily left unsaid by a formalisation. This seems to be an innocuous condition, since expressiveness is generally held by logicians to be a virtuous property of formal logics. If we think of (at least, some) formal logics as formalising types of reasoning, and take their expressive capacity to be a success indicator with respect to their ability to formalise reasoning, then it is not surprising to consider that the mathematical case is analogous. Similar thoughts apply to formal theories of truth, formal epistemology, and so on. Natural language tractability, on the other hand, says that every sentence that can be formulated in $\mathcal{L}_{T_F}$ can be informally parsed as a well-formed, purely mathematical proposition about the subject matter of T_I. That is to say, if T_F formalises a body of informal propositions about some mathematical subject matter, then its language should not produce sentences that evade a natural language treatment. This is not so for formal theories not intended to be formalisations; in that case it is merely desirable that the mathematician should be able to informally express every sentence in the language of the formal theory, for the sake of ease. As Leitgeb explains, at the very least, the terms and formulas of T_F require an interpretation, yet:

> This does not mean that mathematicians *never* take any steps in a [formal] proof on the basis of purely syntactic considerations… but it is essential that it is always possible to switch back into the material [i.e. informal] mode of speaking. (Leitgeb 2009, p. 269)

[6]There may be additional conditions on being a formalisation; namely any condition on being a formal theory is necessarily a condition on being a formalisation, since every formalisation of an informal theory is a formal theory. For example, one might require that a formal mathematical theory, and consequently a formalisation, must be recursively axiomatisable.

In case the condition of natural language tractability appears prohibitively strong to apply to actual mathematical practice, it should be recognised that the practice of formalisation is a *process*, rather than an event, and that informal theorising does not stop once we have one or more available formalisations. If that were the case, mathematics would be a much less rich discipline that it actually is. To return to Lakatos' terms once more, there is not only pre-formal informal theorising, but *post*-formal informal theorising, and there is also formal development of formalisations. A convincing example of how informal theorising and formalisation-development interact might be set theory, where the development of the informal concepts and propositions about sets has proceeded alongside the development and creation of a multitude of formalisations: ZFC, NBG, etc. Formal results (for example, in forcing) have influenced informal theorising (for example, with respect to the continuum hypothesis), and vice versa. In general, T_I can be informed by the results that issue from its formalisation, and vice versa, such that informal theories and their formalisations have a kind of symbiotic relationship. The development of a formalisation may of course give rise to purely formal content that does not have an immediate ancestor in the informal theory T_I with which we started. There are two cases that might arise. Either, this new formal content is afterwards given an informal parsing, which can then be accepted as an extension to T_I, or not. If so, then the new formal content is subsumed under the formalisation T_F of T_I. If not, then it cannot sensibly be said to form part of any formalisation of T_I, by the very meaning of 'formalisation'. It does not formalise anything in T_I. We then might think of there being two formal theories at such a juncture: one is the formalisation T_F of T_I, and the other is a mere formal theory, call it $T_{F'}$, which *contains* T_F. It does not appear too controversial to suppose that a formalisation can cease to be a formalisation once it does not, in its entirety, formalise an informal theory, for in that case the very term 'formalisation' becomes inappropriate to describe that formal theory. The latter—$T_{F'}$—is a formalisation and *then some*.

Not only do we want to ensure that the chosen language of our formal theory is adequate for formalising T_I, we also want to ensure that the class of models that we choose is acceptable. We require a condition on truth preservation because, in the formalisation process from T_I to T_F, we want the right sentences to be given the right truth values in the models of T_F. This is a subtle point: just to say that "a formalisation should preserve truth values" is too strong, since T_I is a *theory* about some part of mathematical reality. Tautologies and contradictions aside, truths about mathematical reality might not be known (or even knowable) with certainty by the T_I-theorist. Yet those T_I-propositions in which the mathematical community has a high degree of belief should have a corresponding formal analogue ϕ, such that ϕ is true in all models of T_F: e.g. 'zero is the smallest natural number'. On the other hand, the existence of a universal set might be considered unsettled by informal theorising, and so its analogous formal sentence is not required to be true, or false, on all models of a given formalisation of set theory. We are left to study the models of T_F so that we might understand better the propositions about T_I's subject matter of whose truth value we are not certain. In sum, truth preservation from T_I to the models of T_F must hold for tautologies, contradictions, axioms, and propositions whose truth values are deemed settled with respect to T_I's subject matter.

Given all of the above, then, we may define a formalisation as follows:

Formalisation

A *formalisation* is a formal theory that is intended to capture an informal theory, and which satisfies: (i) adequate expressiveness; (ii) natural language tractability; and (iii) truth preservation.

It is intended that any proposed formalisation is *faithful* to the informal theory that it is supposed to formalise, and this is ensured to some extent by the three conditions in the above definition. Beyond these necessary properties of a formalisation, though, rival formalisations of the same informal theory can be more or less faithful than each other to that informal theory. For example, one might accept that both PA and PA^2 satisfy all of the necessary conditions upon being a formalisation of arithmetic, yet believe that one of them is a superior formalisation, since it more faithfully captures our informal arithmetical theorising. Faithfulness, beyond the basic level required for formalisation-hood, is measurable in terms of the properties that a formalisation possesses, given the justified beliefs we hold about the informal theory in question. Once we have two bona fide formalisations of T_I, we can look to the beliefs that we hold about T_I's subject matter, and consider which formalisation has the right combination of properties to most faithfully represent those beliefs. Faithfulness, then, is a success term that is relative to the T_I-theorist and her justified beliefs about the subject matter of T_I. It is an informal value-judgement based on formal properties and informal beliefs. We might think, by way of analogy, of language translations: Suppose we commission a translation of an English manuscript into Spanish, and receive two perfectly adequate translations, each of which leaves nothing from the original unsaid, adds no new content, and preserves the meaning of every sentence contained in the manuscript (cf. truth preservation). Still, one translation may be preferable purely for the fact that it is more faithful to the *sense* of the original, due to its choice of idioms, for example. How faithful each translation is to the sense of the original depends on what that sense is taken to be, and different readers of the original may glean different senses from it.

Recognising how the faithfulness of formalisations is measured by formal properties, and its relativity to informal beliefs, is what will allow for a better, more general, evaluation of categoricity's significance. Returning to the discussion of Uniqueness, above, categoricity emerges as a formal property that might be taken to indicate faithfulness to the Uniqueness thesis, in some form or another. Therefore, to make our earlier point a little clearer in light of the above definitions, categoricity is philosophically significant: (i) only with respect to (i.e. as a measure of) the faithfulness of formalisations, not mere formal theories; and (ii) only relative to the particular Uniqueness (and perhaps Existence) beliefs that we hold about the informal theories those formalisations are intended to capture.

17.3 A Framework: Two Kinds of Significance

It is easy to observe that arguments for or against the significance of categoricity fall into two distinct groups. Consider the following claims, for example (emphasis added, unless otherwise stated):

> I conclude that a language *and semantics* [original emphasis] of formalisation should be sufficient to ensure… communication. That is, the language of formalisation should allow categorical characterisations. *It follows that first-order axiomatisations are inadequate*.
> (Shapiro 1985, p. 720)

> It is a property of theories and it is clearly desirable. With a categorical theory… we show that *our axiomatic enterprise has been successful*: we have isolated our intended structure.
> (Meadows 2013, p. 524)

> It is natural, therefore, to think that *a categoricity result will be of use to a realist* concerning a mathematical subject whose language is naturally construed as having an intended interpretation.
> (Walmsley 2002, p. 240)

> Dedekind was well aware that categoricity by itself was worthless… Indeed, I believe that full determinateness of the concept [of natural number] is the only legitimate *justification for the assertion that the concept is instantiated* or that the natural numbers exist.
> (Martin 2012, p. 13)

In these passages, Shapiro and Meadows appear to be concerned with the significance of categoricity insofar as it vindicates a formalisation. Ultimately they disagree over the extent of categoricity's significance, since Meadows argues that upon more careful consideration of categoricity arguments from a novel perspective, it becomes clear that 'we should be reserved before drawing powerful philosophical conclusions' (2013, p. 524) from the property. Conversely, Shapiro, as we shall see in the latter half of this paper, draws some fairly powerful philosophical conclusions from the categoricity of second-order formalisations of certain mathematical theories. Yet, both authors agree that it is, or may be, valuable given a certain kind of informal theory about which we have some antecedent Uniqueness belief(s). Walmsley and Martin instead locate the potential significance of categoricity with the evidence it provides for or against certain philosophical beliefs, presumably given that a particular formalisation is adequate with respect to the informal theory in question. Therefore, we can identify two kinds of philosophical significance for categoricity:

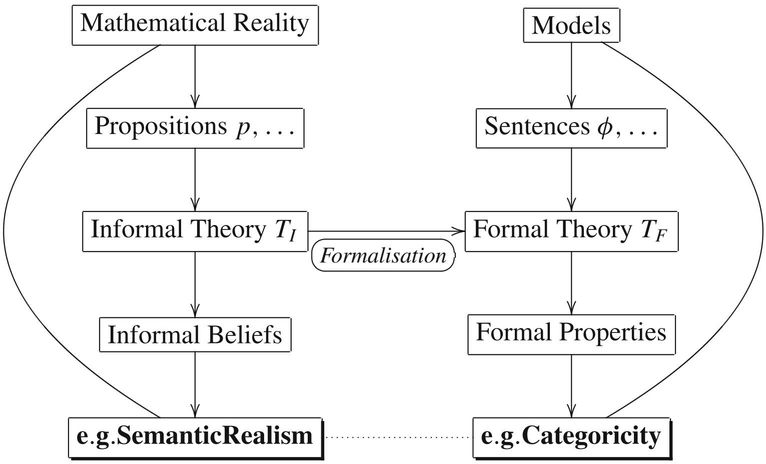
Formal Significance

An argument pertains to the *formal significance* of categoricity if it takes some informal beliefs about T_I as fixed, and assesses the extent to which being categorical makes a formalisation T_F of T_I adequate (i.e. faithful) with respect to those beliefs.

Informal Significance

An argument pertains to the *informal significance* of categoricity if it takes a particular formalisation T_F of T_I as adequate (i.e. faithful), and assesses the extent to which its being categorical (or not) is instructive with respect to what we should informally believe about T_I.

To take a simple example, consider the informal theory of arithmetic. One might claim that categoricity is formally significant because, given that one believes that arithmetic has a unique subject matter—*the* natural numbers—categoricity renders a first-order formalisation of the theory inadequate. The existence of non-standard models in first-order formalisations, by the Löwenheim–Skolem theorem, does not match one's antecedent belief in the uniqueness of the natural numbers. On the other hand, one might claim that categoricity is informally significant because second-order formalisations of arithmetic with full semantics are, for whatever reason,[7] independently justified and thus the fact that they are categorical indicates that we should be realists about the unique structure of the natural numbers. What we have now is a picture that looks like the following:



The parallels between the formal and the informal are not absolute, but where they exist they are fairly obvious. Aspects of mathematical reality, whatever their ontological status, are the truthmakers for propositions that make up informal mathematical theories; meanwhile, models are the truthmakers for sentences that make up formal mathematical theories. Formalisation is what *minimally* hooks up a formal theory—its sentences, models and formal properties—to an informal theory—its propositions, truth makers and our beliefs. As we saw, this minimal hooking up is achieved by checking that a formalisation satisfies adequate expressiveness, natural language tractability and truth preservation. From this basic picture, we can extract a framework for assessing the significance of categoricity and, not only that, but any formal property of a formal mathematical theory. We could even generalise further, to give a framework for assessing the properties of scientific theories and how they affect our theories, models and formalisms on the one hand, and accepted

[7] This would depend on what one takes to be the formal virtues of formal mathematical theories in general. For example, one might have a Quinean preference for first-order formal theories over high-order versions. One could also be more or less committed to various other (perhaps related) properties: expressiveness, ontological parsimony, simplicity, proof-theoretic completeness, etc.

observations on the other. It could also apply to social sciences, economics,...—any theoretical domain where the theorist's aim is to be faithful to some portion of reality.

A formal significance argument assesses categoricity's role in a formalisation's recovery of the independently justified beliefs we have about the informal theory in question. We have been assuming that formalisations are better or worse than each other to the extent that they are *faithful* formalisations, so categoricity's significance for T_F might be that it demonstrate T_F's faithfulness to T_I, given what we believe about T_I. The reverse is true of informal significance arguments. There, we want to know whether T_F's being categorical is necessary and/or sufficient for showing that we were right or wrong all along to believe x, y and z about T_I—on the assumption that T_F is a good formalisation of T_I. In either case, the potential for categoricity's significance can be summed up as somehow aligning our informal beliefs about informal mathematical theories, with the formal properties that result from our formalisation of those theories. That much was already clear in the literature. Yet, the clear admission that either certain informal beliefs are taken as given, or that the adequacy of a particular formalisation is taken as given, is not made in nearly enough arguments in the literature. For example, we saw Walmsley observe that we naturally take a categoricity result for T_F to be useful to the realist about T_I where we hold that there is a unique intended interpretation of T_I. What he is not explicit about, though, is that we must assume that T_F is a good formalisation of T_I in order to really assert this. Otherwise, the link between T_F and T_I is not strong enough for T_F's properties to be definitively informative with respect to the properties we attribute to T_I. For the remainder of the paper, we will take two consequences of this framework in turn.

17.4 Arguing Circularly?

The framework presented above for investigating the philosophical significance of categoricity is backed up by observations on the literature and by the clear methodological value in distinguishing between subtly but substantively distinct arguments. However, we will see shortly that the framework seems to highlight a potential problem for Shapiro's wider project. In particular, we can see from various examples of Shapiro's reasoning about categoricity that one can seemingly argue for *both* formal and informal significance, with respect to a particular informal belief and formalisation. This was not ruled out in the above exposition, but without too much reflection it becomes clear that arguing for both kinds of significance could in some sense be circular.

Formal significance claims seem to assume what informal significance claims try to establish, and vice versa. Claims about the adequacy of a formalisation assume that certain informal beliefs hold, in order to demonstrate this adequacy, or faithfulness to those beliefs. Meanwhile, claims about the justification of certain informal beliefs about a theory's subject matter assume that a certain formalisation is adequate, in order to show how the properties of that formalisation provide evidence for or against that justification. Adequacy or faithfulness of the formalisation is needed in order for it to be worth trusting that formalisation's properties as sources of insight with respect to what we should informally believe. Holding both formal and informal significance claims, with respect to a single formalisation and a single informal belief, simultaneously is essentially to propose claims of the following form:

(a) Given that T_F is a good formalisation of T_I, if T_F is categorical then I was justified in believing b about T_I; and
(b) Given that I was justified in believing b about T_I, if T_F is categorical then T_F is a good formalisation of T_I.

Making both claims with respect to the very same formalisation, and the very same particular informal belief, means assuming what one wants to establish: both the adequacy of the formalisation, and the correctness of the informal belief. Yet, nothing laid out in our framework forbids a philosopher from arguing for both the formal and informal significance of categoricity. Shapiro even seems to do just this, as we will see. At this point, then, it looks as though interpreting Shapiro's two major projects—structuralism and second-order logic—within the present framework makes those projects seem in danger of being uninformative or circular with respect to the role of categoricity. We will explore Shapiro's two significance claims in more detail before considering a number of responses to the circularity that seems to threaten from taking these claims together.

17.4.1 Shapiro's Two Significance Claims

In support of second order logic, we saw at the start of Sect. 17.3 that Shapiro deems 'first order axiomatisations… inadequate' by their not being categorical and thus not formally recovering the property of communicability, with respect to 'non-algebraic'[8] mathematical theories.[9] This reasoning is clearly focused on the success of formalisations at doing what we manage to do informally; in other words, at formally recovering the Uniqueness property that we informally—whether by our philosophical reasoning or merely linguistically—attribute to the subject matters of arithmetic and other mathematical theories. Categoricity is taken here by Shapiro to adjudicate between first- and second-order formalisations of arithmetic. This is one of the clearest examples in the literature of a formal significance claim for categoricity. The claim, simply put, is that given that the obvious communicability of e.g. arithmetic implies at least one version of the Uniqueness thesis, PA^2 with full semantics is a good formalisation of the theory, because it is categorical.

[8]These are theories which seem to have an intended interpretation.

[9]Shapiro believes in communicability to the extent that he thinks it is observable and in need of explanation.

However, Shapiro could also be interpreted as arguing for the informal significance of categoricity, as a way of bolstering his structuralist theses about the subject matter of non-algebraic theories. Indeed, Horsten (2007) observes that 'the appeal to second-order logic appears as the final step in the structuralist project of isolating intended models of mathematics', via categoricity. Shapiro's informal beliefs are summed up by the following statement:

> The view in question is what I call *ante rem* structuralism, the thesis that mathematical structures exist prior to, and independent of, any exemplifications they may have in the non-mathematical world.
>
> …According to ante rem structuralism, natural numbers are places in structures, and places in structures are bona fide objects… numerals are singular terms.
> (2006, pp. 109–10)

So, we can see that Shapiro's background informal beliefs about (at least) arithmetic are ontological realism and semantic realism. He also states the following:

> It is fair to say that when it comes to mathematics and theories of other *abstracta*, realism in ontology often falters over reference (about as much as it falters over epistemology).
>
> …If we stick to categorical (or semantically complete) theories, realism in ontology, as construed here, thus leads to realism in truth-value.
> (1997, pp. 139–40)

Moreover:

> [The second-order logic advocate] takes the categoricity results to *confirm* his prior belief that arithmetic is unambiguous. [Original emphasis.]
> (1991, p. 207)

It is not very suspect to infer that Shapiro *himself* takes categoricity to 'confirm' his prior belief that arithmetic is unambiguous, since he is explicitly a second-order logic advocate. Neither is it illegitimate to propose that by 'unambiguous' Shapiro intends at least one Uniqueness thesis, and that given the above remarks one of these is likely to be semantic realism about arithmetic. It is not too much of a liberty, therefore, to interpret an informal significance claim in Shapiro's work, alongside the more explicit formal significance claim that he makes for categoricity. The claim is that given that PA^2 with full semantics is a good formalisation of arithmetic and it is categorical, the belief in semantic realism about arithmetic is confirmed. Yet, that PA^2 with full semantics is the preferable formalisation of arithmetic is precisely the claim Shapiro aims to establish in (1991), via (in part) its categoricity, given some realist beliefs about arithmetic that are based on some observations on mathematical practice.

If he is genuinely attributing both formal and informal significance to categoricity, then Shapiro's structuralism and second order logic projects may be uninformative in tandem with one another, at least to the extent that categoricity is involved with respect to a particular formalisation and a particular informal belief. We have seen that under one interpretation, he appears to argue for the adequacy of PA^2 with full semantics as a formalisation of arithmetic, on the basis of semantic realism, and for the justification of semantic realism on the basis of the adequacy and categoricity of

PA^2 with full semantics. One might conclude that either, one of his arguments for the significance of categoricity needs to be removed such that either structuralism or second order logic has one less piece of support, or we need evidence that this worry can be assuaged. More generally, if this stands as a formidable worry, then it limits the potential significance of categoricity by *half*, for any individual philosopher or philosophy of mathematics: perhaps the property can only be attributed either formal or informal significance, but not both.

17.4.2 Possible Responses

We all know that there are bad, and not so bad, kinds of circularity. It is worth asking, therefore, how worrying the above result is for Shapiro—whether it can be shown to be an acceptable kind of circularity, or whether it can be avoided altogether. We will look at four possible responses. The first claims that our analysis of significance is inappropriate, but this option is shown to be against the present spirit of clarity. The second is a more sophisticated version of the first response, arguing that Shapiro could escape circularity if one of his significance claims is general and the other is specific. It is argued that such a defence would rely on a general claim not implying a specific instance that reintroduces a circle when taken with the other, specific, claim. This is not easy to do without deferring to the third response considered: that Shapiro's structuralism and second-order logic theses are in reflective equilibrium. This reply would deem our significance framework inapplicable to Shapiro's reasoning, since it does not take reflective equilibrium between formal and informal significance claims into account. However, textual evidence is presented that seems to suggest that Shapiro could not consistently make this defence, since he rejects Resnik's use of reflective equilibrium when claiming that logic is not objective, and since Shapiro clearly deems mathematics to be objective. The final response suggests that Shapiro could retreat to naturalism, or working realism, to circumvent the circle and still establish justification for both structuralism and second-order logic via categoricity. This response will also not do, for these retreats threaten to collapse quickly back into straightforward formal and informal significance claims, reinstating the suspected circularity.

17.4.2.1 Square Pegs and Round Holes

Suppose we take an even more general perspective on the whole debate over the philosophical significance, not just of categoricity (which is a special case), but of any formal mathematical property. We can think of the debate as arising out of a desire to secure and justify mathematical philosophy in general. In 'mathematical philosophy', we take informal beliefs and arguments, and formalise them in case increased precision might illuminate the faults, ramifications and so on of our pre-formal thought; but there is a gap between the informal and the formal that we must be sure to have closed when we do mathematical philosophy. From this most general perspective, one could argue that the significance of categoricity is as a ‘bridging property’—a property that secures the process of formalisation itself. The key shift in perspective is that such a property is significant, not in any particular *direction*, either towards justifying a formalisation, or towards justifying an informal belief, but to the extent that it underpins the idea, in general, that the informal *can* be formalised. A bridge has no direction, but merely connects two places (i.e. the formal and the informal). According to this view, perhaps categoricity’s significance does not rest on taking either informal mathematical beliefs, or formalisations, as independently justified, and so interpreting Shapiro’s claims according to our framework is not legitimate. A directional framework is not apt for analysing a directionless feature. You might say: ‘square pegs and round holes’.

This approach, however, seems to re-blur the lines of the debate. We have already seen in examples from the literature that there is good reason to think there are two distinct, directional ways of arguing for or against the philosophical significance of categoricity. This was the evidential basis for the present framework. Moreover, it is clear that delineating between existing discussions of subtly, yet substantively, distinct matters is methodologically sound and fruitful. This is the rational basis for our framework. It seems, therefore, that we should be reluctant to throw it away and head back towards imprecision and/or implicit assumptions on the basis that some existing philosophy is rendered potentially problematic by it. In the response above, however, the underlying beliefs of the arguer are in danger of being pushed out of focus once more. For that reason, it is not a satisfying response to our problem. The next response builds on this initial attempt at recovery to offer a more sophisticated answer.

17.4.2.2 Differentiating Between General and Specific Claims

Another consideration[10] that may allow Shapiro to escape potential circularity in his wider project is that we might distinguish between general and specific significance claims. The case we have had in mind for Shapiro seems to be of two *specific* claims about categoricity’s significance, with respect to arithmetic. That is, we have one specific formal claim in favour of PA^2 and one specific informal claim in favour of semantic realism about the natural number structure. But perhaps there is room to avoid circularity by making a general claim in one direction and a specific claim in the other. For example, Shapiro might start out from a preference for second-order formal theories *generally* and let their categoricity guide him towards Uniqueness beliefs about informal theories. This is a general informal significance claim. Then, having honed those Uniqueness beliefs into particular theses on the basis of other philosophical considerations, make the specific formal significance claim that semantic realism about arithmetic vindicates, in particular, PA^2.

[10] I am indebted to an anonymous referee for suggesting this response to me.

On the face of it, this seems to represent a reasonable response to the circularity problem. However, a potential trap should be underlined. Typically, when a general claim is made, all specific instances follow from it immediately. This means that the general significance claim that is made in one direction may imply a specific significance claim in the same direction, which is itself the mirror of the specific significance claim in the opposite direction, which the theorist wants to make. If that were the case, circularity between two specific claims would be merely hiding as an implication of a formal and an informal significance claim—one of which is general. Shapiro would have to convince us that, for example, a starting preference for second-order theories in general does not just immediately imply a prior preference for PA^2 specifically, which would close the gap just cleaved between general and specific significance claims and reinstate a circle. This is the kind of careful footwork that is suggested by proponents of the following response, which is a further refinement of the response just discussed. As we will see, this next response is not easy to reconcile with claims Shapiro has made elsewhere about mathematics and logic and, as such, if the distinction between general and specific significance claims is to work, Shapiro must convince us that general claims do not imply certain specific ones, without relying on the following response. That would by all means be an achievable task, but one that would require further philosophical work.

17.4.2.3 Reflective Equilibrium

Goodman gives the following description of what is known as 'reflective equilibrium', which he takes to dissolve the ('old') problem of induction:

> I have said that deductive inferences are justified by their conformity to valid general rules, and that general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.
> (1983, p. 64)

Replace 'rules and accepted inferences' with 'accepted formal virtues of formalisations, and informal philosophical beliefs', and this represents a potential defence for Shapiro. The idea would be that we are permitted to move between beliefs and theoretical considerations in a justification process towards a balance of coherent judgements. Specifically, the mathematician can start out with some informal beliefs about the natural numbers, say, and have them inform her choice of formalisation of arithmetic, and then having considered the consequences of that formalisation, use those properties to further confirm (or amend) her prior beliefs. Shapiro can start out a semantic realist about arithmetic and allow this to set his preference for a categorical formalisation, and then seeing the success of such a formalisation, use its categoricity to further affirm his belief in semantic realism. The outcome, on this view, is not circularity of any vicious sort, but simply a pair of theses justified within a coherentist-style picture. Encouragingly, this reasoning seems to fit roughly within Shapiro's epistemology, which rejects the 'foundations' of foundationalism. Furthermore, Williamson (2007, p. 244) draws a parallel between reflective equilibrium arguments in philosophy, and 'the mutual adjustment of theory and observation in natural science'. This lends further support to the applicability of reflective equilibrium to Shapiro's reasoning, since we have already claimed that much of what is said about formal and informal theorising in mathematics applies equally to scientific theories and observations on physical reality.

However, there are issues with the reflective equilibrium defence. The most obvious is Shapiro's own objection to Resnik's support for reflective equilibrium as a methodology for the logician—a case that has obvious relevance to the mathematical case we are considering. Resnik claims:

> One begins with one's own intuitions concerning logical correctness (or logical necessity)… One then tries to build a logical theory whose pronouncements accord with one's initial considered judgements. It is unlikely that initial attempts will produce an exact fit between the theory and the 'data'… Moreover, in deciding what must give, not only should one consider the merits of the logical theory *per se*… but one should also consider how the theory and one's intuitions cohere with one's other beliefs and commitments, including philosophical ones. When the theory rejects no examples one is determined to preserve and countenances none one is determined to reject, then the theory and its terminal set of considered judgements are in… *wide reflective equilibrium*. Resnik (1997, p. 159)

Shapiro challenges Resnik's account of logic, claiming that 'logic is *objective* if anything is' (2000, p. 335) and that if Resnik's reflective equilibrium account of the methodology of logic is correct, the discipline's 'objectivity is brought into doubt, at least potentially' (2007, p. 363). Shapiro is also happy to 'confess to the… intuition that mathematics is objective' (2007, p. 338). If we can show that Shapiro must reject his intuition about the objectivity of mathematics in order to use a reflective equilibrium defence against the circularity problem, that defence may be unpalatable. The threat to objectivity from reflective equilibrium, if it is true, does seem to generalise from logic to mathematics. For Shapiro, reflective equilibrium rules out logical pluralism from the point of view of the logical theorist, and insulates her from objectivity. The same, however, can be said in the mathematics case: nothing privileges Alex's reflective equilibrium between arithmetic, PA^2 and categoricity's considerable significance, over Ben's reflective equilibrium between arithmetic, PA and categoricity's lack of significance. From Shapiro's two premises—the objectivity of mathematics, and the claims that reflective equilibrium sacrifices the objectivity of logic—we can argue that, for Shapiro, reflective equilibrium must also sacrifice the objectivity of mathematics. This would make a reflective equilibrium response to the circularity worry over his two significance claims regarding categoricity untenable with his claim that mathematics is objective.

So, it looks as though Shapiro at least cannot retreat to a reflective equilibrium defence *easily*. He would either have to renounce his claim that reflective equilibrium compromises the objectivity of logic, or that mathematics is objective, or otherwise provide a reason to reject the inference from his objection to Resnik, to the claim that he must reject reflective equilibrium in the mathematical case too. Shapiro does admit that his intuition over mathematics' objectivity 'may be little more than a prejudice, subject to correction in light of philosophical theorizing (or incredulous stares) and further articulation of the notion of objectivity' (2007, p. 338). In other words, it is open to Shapiro to reject the above reasoning on the basis that his belief in mathematics' objectivity is no more than naive intuition, whilst his belief in the objectivity of logic has substantial and reasoned justification. If he lets go of mathematics' objectivity, then he can consistently claim that his two projects are in reflective equilibrium and therefore not subject to circular reasoning regarding the significance of categoricity. Yet, despite Shapiro's modest admission that his claim rests on intuition, he also concludes that 'mathematics comes out objective on every one' of Wright's (1992) tests, or criteria, for objectivity. Moreover, he explicitly states that the 'objectivity of logic is tied to other discourses, and vice versa… if logic fails to be objective in some sense and to some extent, then mathematics fails to be objective in the same sense, to the same extent' (2007, p. 365). None of this is to say that any of these claims are correct, but that by holding them, Shapiro has a more difficult task of using reflective equilibrium to avoid circularity. Even a simple retreat to a methodology that apparently fits rather well with Shapiro's epistemology of mathematics is not so simple after all. One of his previous claims has got to give.

17.4.2.4 A Retreat to Naturalism

A final response to consider is that we may not have taken Shapiro's background philosophy of mathematics seriously enough. We might reasonably call Shapiro a 'naturalist', or 'mathematics first' proponent. According to this view, philosophy of mathematics must adhere to a 'faithfulness constraint' (see Shapiro 2006) with respect to actual mathematical practice, i.e. the reasoning employed should always take into account the mathematician and her theorising. Note that this does not immediately undermine Shapiro's claims towards the formal significance of categoricity, since the informal beliefs he thinks second-order formalisations successfully remain faithful to are themselves directly informed by mathematical practice. For example, he observes that we seem to successfully communicate with one another about 'the natural numbers', 'the real numbers', and so on, in practice (Shapiro 1991, p. 102). This leads to an informal belief in communicability, which Shapiro thinks can be restored with second- rather than first-order theories, by the fact that categoricity rules out misunderstandings between interlocutors on the basis of non-standard models.[11] And of course, he takes ante rem structuralism to be perfectly in line with the faithfulness constraint and naturalism in general.

However, this consideration itself is problematic in two related ways, both of which show that the circularity worry may in fact be inherent in his starting assumptions. First of all, we may have arrived at the justification of semantic realism via the correctness of second-order formalisations plus naturalism, yet *still* via the informal

[11] See Parsons (1990) and McGee (1997) for alternative ways of ruling out misunderstandings issuing from non-standard models.

significance of categoricity in a roundabout way. The naturalist can wave away the worry expressed by Väänänen (2001, p. 509), that 'we can only speculate whether our formalisation captures what we intend' by presuming that we do not intend formalisations to capture anything. Presumably, for the naturalist, the formalisation is given by the mathematician and we philosophers are to investigate what they capture and imply, by first taking the mathematician's language and practice at face value. Yet, since naturalism is now doing the work of an informal significance claim, on behalf of semantic realism and on the back of second-order formalisations, we can ask whether naturalism itself is uninformatively circular when taken together with a formal significance claim in favour of second-order formalisations. In basic terms, using naturalism as a defence of Shapiro's twin projects does nothing more than replace the argument for the informal significance of categoricity with the naturalist assumption that the mathematician's use of '*the* natural numbers' implies semantic realism about the natural numbers. Basically, the naturalist takes categoricity just as seriously as the philosopher who wants to claim informal significance for it—both believe that categoricity is a source of philosophical belief.

The second, related worry takes this a step further. Shapiro also describes himself in (1991) as a 'working realist': someone who believes that mathematical practice should proceed as if realist beliefs about mathematical theories were true. This is not very far away from assuming what formal significance claims seek to establish, analogously to the case of naturalism and informal significance claims. Working realism has mathematical practice, which presumably includes formalisation choice, proceeding as if (i.e. *given that*) some informal realist beliefs are true. It does not make much of a difference that a formal significance claim is based on informal beliefs being true whilst working realism is a little less committed, only recommending that we theorise 'as if' those beliefs were true. Shapiro would contend, no doubt, that working realism is not a substantive realist thesis, and is compatible with anti-realist informal beliefs about ontology and suchlike, but for all practical purposes all we *can* do is theorise as if our informal beliefs are true, if we have justification for them. Surely, he would not recommend adopting working realism on the basis of unjustified realist beliefs about mathematical theories. It seems likely that Shapiro would also agree that we cannot outright know whether or not those beliefs are true. Therefore, the difference between an assumption of working realism and a proposal of informal significance for categoricity is not sharp enough. In sum, the problem is effectively pushed back a level: naturalism and working realism taken together are (at least close to being) uninformatively circular, and trying to evade the circularity of simultaneous informal and formal significance claims by replacing one or the other with the corresponding assumption of naturalism or working realism, respectively, does not get rid of the circularity. Linnebo asserts a very similar claim:

> Shapiro denies that philosophy-last [or, 'mathematics first'] implies *philosophical* realism, because this form of realism aims to give an extra-mathematical account of mathematical practice. I am not entirely convinced. It seems to me a real danger that philosophy-last *will* imply philosophical realism and thus obliterate any distinction between Shapiro's two forms of realism… This leaves his central distinction between working realism and philosophical realism… in danger of collapsing. Linnebo (2003, pp. 94–95)

If Linnebo is right, then the uninformative circularity problem that we have raised for Shapiro's philosophy of mathematics is in fact just a special case of a larger problem with the incompatibility of assuming working realism and a naturalist approach from the outset. That wider problem is beyond our scope, but if it stands, then it adds weight to our criticism of Shapiro's co-existing formal and informal significance arguments. Their uninformativeness becomes predictable. It is worth noting that Shapiro himself admits that he has 'nothing to offer' (1997, p. 48) with respect to the wider problem that Linnebo points out. Without being too uncharitable, one might suspect that the same would hold for our specific case in relation to categoricity.

In summary of the first consequence for our framework for assessing the significance of categoricity, we have seen that there is a circularity threat when a philosopher makes both a claim for the formal significance of categoricity, and a claim for the informal significance of categoricity. A specific instance of this dual claim in the literature, via some interpretation, is Shapiro's use of categoricity to support both informal (e.g. semantic realist) beliefs about the subject matters of certain informal theories, and formalisations of those theories (with second-order logic and full semantics). Four responses were considered, although there are doubtlessly more available. No knock-down objection was given to any of these options. Yet, by inspecting them in some detail the need for more exposition of background beliefs and assumptions in any approach towards the significance or otherwise of categoricity is needed. In any case, none of these responses provide an easy retreat for Shapiro from the threat of circularity; if anything we were led to consider a more general version of the problem, identified by Linnebo, that may strengthen the credibility of our circularity claim.

17.5 The Limited Significance of Categoricity

Our final goal is to show that categoricity has limited formal and informal significance with respect to informal beliefs about informal mathematical theories, with a case study. The semantic realist about T_I believes that every purely mathematical, natural language proposition about T_I's subject matter has a determinate truth value. The semantic anti-realist denies this. As such, the categoricity of a given formalisation T_F of T_I is *formally* significant to the extent that it is necessary and/or sufficient for T_F's accurately reflecting whichever belief is independently justified: semantic realism, or semantic anti-realism, about T_I's subject matter. The categoricity of T_F is *informally* significant to the extent that it is necessary and/or sufficient for telling us whether to be semantic realists or semantic anti-realists about the subject matter of T_I, given that T_F is a good formalisation of T_I. Suppose that the formal analogue of every T_I-proposition having a determinate truth value is that every T_F-sentence has the same truth value in all models. This formal property is 'elementary equivalence': the models of a formal theory are elementary equivalent if they make all the same sentences true, and all the same sentences false. It is not difficult to see why one might posit a conceptual link between semantic realism and elementary equivalence.

If semantic realism about T_I is justified, therefore, then we have good reason to expect that a formalisation T_F of T_I is adequate only if its models are elementary equivalent; otherwise we have good reason to think that T_F is inadequate. If semantic anti-realism about T_I is justified, then an adequate formalisation should not have elementary equivalent models, and a formalisation with elementary equivalent models should be deemed inadequate. Likewise, if T_F is a good formalisation of T_I, then we have good reason to be semantic realists if T_F's models are elementary equivalent; otherwise we have good reason to be semantic anti-realists. The question for our case study then becomes: is the categoricity of T_F necessary and/or sufficient for T_F's having elementary equivalent models?

In fact, it is easy to show that a categorical formal theory must have elementary equivalent models, but that a formal theory with elementary equivalent models need not be categorical. Namely, there are first-order formal theories with elementary equivalent models, and as we know, no first-order formal theory with an infinite domain can be categorical. If we accept that elementary equivalence of T_F's models is the formal evidence for semantic realism about T_I, then what this says is that categoricity is only sufficient, and not necessary, for having good reason to accept T_F as a good formalisation of T_I, given an antecedent belief in semantic realism about T_I. Sufficing for elementary equivalence is not enough for categoricity to have any substantial kind of significance for the semantic realist. Suppose that one assumes semantic realism about arithmetic and wonders what is the best formalisation of the theory. Since categoricity is only sufficient and not necessary for elementary equivalence, we have no *definitive* way of choosing between a categorical and a non-categorical formalisation; the semantic realist is not even interested in this, as long as the formalisation in question has elementary equivalent models. The semantic realist might as well deal only with elementary equivalence and eschew talk of categoricity's significance. If the semantic realist looking for a good formalisation can get by in their investigation without resorting to categoricity, then it cannot be significant for them. Exactly analogous remarks apply to the T_F-ist who would like to know what T_F tells them they should believe about truth values in T_I. Of course, categoricity can be methodologically useful in pointing these philosophers towards elementary equivalence, but it cannot be said to do substantial *philosophical* or 'conceptual' work for them. It is a mere stepping stone: significant, but not in a huge way.

17.6 Conclusion

The main aim of this paper was to get maximally clear on how to assess the significance of categoricity, in order to encourage increased precision and explicitness about background assumptions. It was argued that there are two kinds of significance that categoricity could have, and that arguing for both kinds simultaneously may be uninformatively circular, at least in Shapiro's case given his other background beliefs. It was also shown that, on the basis of the framework, it turns out that categoricity has limited significance of either kind for the semantic realist.

References

Antonutti Marfori, M. (2010). Informal proofs and mathematical rigour. *Studia Logica*, *96*, 261–272.

Button, T., & Walsh, S. (2015). Ideas and results in model theory: Reference, realism, structure and categoricity. Draft.

Corcoran, J. (1980). Categoricity. *History and Philosophy of Logic*, *1*, 187–207.

Dedekind, R. (1888). *Was sind und was sollen die Zahlen*. Braunschweig: Viewig.

Ferreirós, J. (2011). On arbitrary sets and ZFC. *The Bulletin of Symbolic Logic*, *17*(3), 361–393.

Gaifman, H. (2004). Non-standard models in a broader perspective. In A. Enayat & R. Kossak (Eds.), *Nonstandard models of arithmetic and set theory: AMS special session, January 15–16 2003* (Vol. 361, pp. 1–22). Contemporary Mathematics. Providence: American Mathematical Society.

Gödel, K. (1951). Some basic theorems on the foundations of mathematics and their implications. In S. Feferman et al., (Eds.), *Kurt Gödel: Collected works. Unpublished essays and letters* (Vol. 3, pp. 304–323). Oxford: Oxford University Press.

Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge: Harvard University Press.

Halbach, V., & Horsten, L. (2005). Computational structuralism. *Philosophia Mathematica*, *13*, 174–186.

Hellman, G. (1989). *Mathematics without numbers: Towards a modal-structural interpretation*. Oxford: Clarendon Press.

Hintikka, J. (2011). What is the axiomatic method? *Synthese*, *183*, 69–85.

Horsten, L. (2007). Philosophy of mathematics. *Stanford Encyclopedia of Philosophy*.

Isaacson, D. (2011). The reality of mathematics and the case of set theory. In Z. Novak & A. Simonyi (Eds.), *Truth, reference, and realism* (pp. 1–76). New York: Central European University Press.

Kripke, S. (1976). Is there a problem about substitutional quantification? In G. Evans & J. McDowell (Eds.), *Truth and meaning* (pp. 324–419). Oxford: Oxford University Press.

Lakatos, I. (1978). What does a mathematical proof prove? In J. Worrall & G. Currie (Eds.), *Mathematics, science and epistemology: Philosophical papers* (Vol. 2, pp. 61–69). Cambridge: Cambridge University Press.

Leitgeb, H. (2009). On formal and informal provability. In Ø. Linnebo & O. Bueno (Eds.), *New waves in philosophy of mathematics* (pp. 263–299). Palgrave Macmillan.

Linnebo, Ø. (2003). Review of Philosophy of mathematics: structure and ontology. *Philosophia Mathematica, 11*, 92–104.

Martin, D. A. (2001). Multiple universes of sets and indeterminate truth values. *Topoi*, *20*, 5–16.

Martin, D. A. (2012). Completeness or incompleteness of basic mathematical concepts. Draft.

McGee, V. (1997). How we learn mathematical language. *Philosophical Review*, *106*, 35–68.

Meadows, T. (2013). What can a categoricity theorem tell us? *The Review of Symbolic Logic*.

Melia, J. (1995). On the significance of non-standard models. *Analysis*, *55*, 127–134.

Parsons, C. (1990). The uniqueness of the natural numbers. *Iyuun*, *13*, 13–44.

Resnik, M. (1997). *Mathematics as a science of patterns*. Oxford: Oxford University Press.

Shapiro, S. (1985). Second-order languages and mathematical practice. *The Journal of Symbolic Logic*, *50*(3), 714–742.

Shapiro, S. (1991). *Foundations without foundationalism*. Oxford: Oxford University Press.

Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. Oxford: Oxford University Press.

Shapiro, S. (2000). The status of logic. In P. Boghossian & C. Peacocke (Eds.), *New essays on the a priori* (pp. 333–366). Oxford: Oxford University Press.

Shapiro, S. (2006). Structure and identity. In F. MacBride (Ed.), *Identity and modality* (pp. 109–145). Oxford: Oxford University Press.

Shapiro, S. (2007). The objectivity of mathematics. *Synthese*, *156*, 337–381.

Väänänen, J. (2001). Second-order logic and foundation of mathematics. *The Bulletin of Symbolic Logic*, *7*(4), 504–520.

Walmsley, J. (2002). Categoricity and indefinite extensibility. *Proceedings of the Aristotelian Society, New Series*, *102*, 217–235.
Wang, H. (1955). On formalization. *Mind*, *64*(254), 226–238.
Welch, P. (2012). Conceptual realism: Sets and classes. Draft.
Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.
Wright, C. (1992). *Truth and objectivity*. MA: Harvard University Press.