Boston Studies in the Philosophy and History of Science 339

Gianluigi Oliveri
Claudio Ternullo
Stefano Boscolo *Editors*

Objects, Structures, and Logics

FilMat Studies in the Philosophy of Mathematics

Springer

Boston Studies in the Philosophy and History of Science

Volume 339

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 https://link.springer.com/bookseries/5710

Gianluigi Oliveri • Claudio Ternullo
Stefano Boscolo
Editors

Objects, Structures, and Logics

FilMat Studies in the Philosophy of Mathematics

Springer

Editors
Gianluigi Oliveri
Accademia Nazionale delle Scienze, Lettere ed Arti di Palermo
Università di Palermo
Palermo, Italy

Claudio Ternullo
Departament de Matemàtiques i Informàtica
Universitat de Barcelona
Barcelona, Spain

Stefano Boscolo
Dipartimento di Filosofia e Beni Culturali
Università di Venezia Ca' Foscari
Trevio, Italy

ISSN 0068-0346 ISSN 2214-7942 (electronic)
Boston Studies in the Philosophy and History of Science
ISBN 978-3-030-84705-0 ISBN 978-3-030-84706-7 (eBook)
https://doi.org/10.1007/978-3-030-84706-7

Contents

About the Editors

Claudio Ternullo (PhD Liverpool, 2012) is currently Beatriu de Pinós (Marie-Skłodowska Curie Actions COFUND) Postdoctoral Fellow at the University of Barcelona. Previously, he has held positions as post-doc at the Kurt Gödel Research Center for Mathematical Logic at the University of Vienna and at the University of Tartu. His research interests lie in logic and the philosophy of mathematics, in particular in the philosophy of set theory. His work focuses on the set-theoretic multiverse, new axioms (and their justification), and mathematical platonism (in particular, Gödel's Platonism). He has also done work on issues in ancient and medieval philosophy.

Stefano Boscolo (PhD Palermo, 2016) is an IT Solution Architect at Volkswagen Group. His current areas of expertise are advanced analytics, cloud computing, and machine learning. After receiving his PhD in Logic and Philosophy of Science, he worked at Ca' Foscari University of Venice on truth pluralism. Throughout his academic career, he worked on the philosophy of mathematics, in particular on the debate between platonism and anti-platonism. His current research interests range from natural-language processing to deep neural networks.

Gianluigi Oliveri obtained a *laurea* (BA) in philosophy from the University of Bari (Italy). After that, he received a DPhil in philosophy from the University of Oxford (GB) and a PhD in cognitive science from the University of Messina (Italy).
His research interests range from the philosophy of mathematics, artificial intelligence, and the philosophy of science to metaphysics and the history of philosophy.

He has held teaching and research posts at the universities of Reading, Leeds, Keele, Oxford, and Palermo, and has been a visiting scholar at the Sydney Centre for the Foundations of Science, University of Sydney, Australia. He is, currently, Associate Professor of Logic and the Philosophy of Science at the University of

Palermo (Italy); corresponding member of the Accademia Nazionale di Scienze, Lettere ed Arti di Palermo (National Academy of Sciences, Letters and Arts of Palermo); and chairman of the Centro Interdipartimentale per le Tecnologie della Conoscenza (Interdepartmental Centre for the Technologies of Knowledge) at the University of Palermo.

Contributors

N. Barton Fachbereich Philosophie, University of Konstanz, Konstanz, Germany

S. Bianchi University School for Advanced Studies (IUSS), Pavia, Italy

I. Blechschmidt Institut für Mathematik, Universität Augsburg, Augsburg, Germany

A. Cantini DILEF, Università di Firenze, Firenze, Italy

O. Caramello Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Como, Italy
Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France

D. Catta LIRMM – Montpellier University, Montpellier, France

L. Crosilla Department of Philosophy, IFIKK, University of Oslo, Blindern, Norway

A. P. d'Aragona Centre Gilles Gaston Granger, Aix-Marseille Univ, CNRS, Aix-en-Provence, France

J.-P. Marquis Département de Philosophie, Université de Montréal, Montréal, QC, Canada

S. Maschio Dipartimento di Matematica "Tullio Levi-Civita", Padova, Italy

E. Moriconi Department of Civilisations and Forms of Knowledge, via P. Paoli 15, Pisa, Italia

R. Samuels Ohio State University, Columbus, OH, USA

S. Shapiro Ohio State University, Columbus, OH, USA

E. Snyder LMU, Munich, Germany

D. Waxman Department of Philosophy, National University of Singapore, Singapore

L. Zanetti NEtS Center, Department of Humanities and Life Sciences, Scuola Universitaria Superiore IUSS Pavia, Pavia, Italy

Chapter 1
Introduction

Gianluigi Oliveri, Claudio Ternullo, and Stefano Boscolo

1.1 A Metaphysical Dispute

A very important controversy raging within present-day philosophy of mathematics is the so-called 'realism/anti-realism dispute.' In such a dispute the realist, who holds that mathematics is about discovering and describing properties of entities which exist independently of our knowledge, is opposed by the anti-realist who does not share his conviction.

Within the analytical tradition, the realism/anti-realism dispute about mathematics is considered to be metaphysical, because the belief in the existence (or non-existence) of mathematical reality is an essential component of our thought about the world.[1] To this, we must add that if a belief X is an essential component of our thought about the world then X is one of the preconditions of experience, not an empirical hypothesis, that is, a belief the correctness of which can be empirically controlled.

[1] See on this Strawson (1974), Introduction, p. 9.

G. Oliveri (✉)
Accademia Nazionale delle Scienze, Lettere ed Arti di Palermo, Università di Palermo, Palermo, Italy
e-mail: gianluigi.oliveri@unipa.it

C. Ternullo
Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain
e-mail: claudio.ternullo@ub.edu

S. Boscolo
Dipartimento di Filosofia e Beni Culturali, Università di Venezia Ca' Foscari, Trevio, Italy
e-mail: stefano.boscolo@unive.it

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_1

Indeed, if we consider our best scientific theories, that is, theory of relativity and quantum mechanics, it is undeniable that they are part of our thought about the world and that some mathematical theories are not only applicable, but also indispensable, to them. Therefore, the belief in the existence/non-existence of the posits of those mathematical theories which are indispensable to theory of relativity or to quantum mechanics is an essential component of our thought about the world; and a precondition of our experience, for how this is possible through theory of relativity and quantum mechanics. But, it is not one of the hypotheses that theory of relativity and quantum mechanics strive to confirm/disconfirm experimentally.

1.2 Can We Dispense with Metaphysics?

Some authors have argued that metaphysical questions about mathematics of the sort we have raised at the beginning of the previous section can be 'put to one side,' if we are prepared to accept the view that mathematical activity simply consists in devising lists of axioms to do various important mathematical jobs, and criticising or defending such axioms in relation to their effectiveness towards these ends.[2]

In replying to this, we, preliminarily, observe that the realism/anti-realism dispute in the philosophy of mathematics concerns the subject matter of mathematics, not what mathematicians and, in particular, set theorists do in their work.

Secondly, asserting that developing mathematical theories consists in devising a list of axioms to do various important mathematical jobs, etc. not only is false in general—as, for instance, the history of number theory and of mathematical analysis clearly show—but, in particular, it runs against the history of set theory. As is well known, when Cantor first introduced, and then developed, set theory, he proceeded in an entirely informal way. In fact, the first sustained attempt to axiomatise set theory was made by Zermelo only in 1908.[3]

Lastly, assuming, for the sake of argument, that, of all mathematical theories, set theory could be simply reduced to devising a list of axioms to do various important mathematical jobs, etc., we would then have to ask ourselves how these axioms are chosen. In answering such a question, we would have to say that either they are, like the rules of a game, mere tools invented to produce a certain result; or, alternatively, we would have to give an account of their non conventional nature. It is not difficult to see that both these ways of explaining the process of axiom choice in set theory are bound to come shrouded in a thick cloud of metaphysics reintroducing, as it were, from the window what had been forcibly pushed out of the door: the realism/anti-realism dispute about set theory.

[2] Restricted to set theory, this is the position of P. Maddy. Those interested in Maddy's philosophy of set theory might find useful to read Maddy (2011).

[3] See Zermelo (1908).

1.3 An Ontological Dispute

A realist philosopher of mathematics, besides arguing in favour of the existence of mathematical reality, is under the obligation to provide an account of what sort of entities populate mathematical reality, e.g., abstract objects, abstract structures, mental phenomena, attributes of sensible objects, etc. And, although the anti-realist is spared some of the chores the realist has to go through, such as compiling a list of what sort of things constitute mathematical reality, etc., he must, nevertheless, account for mathematical activity by saying what this consists in, e.g., game invention and playing, production of formal constructions, invention of linguistic tools for the articulation of the language of theories belonging to the empirical sciences, etc. In any case, the answers to the unavoidable questions 'What kind of things populate mathematical reality?' and 'What kind of things embody mathematical activity?' are going to contribute to a debate concerning the nature (ontology) of mathematical reality/activity. And this is a debate which radically differs from the realism/anti-realism dispute.

To see this more clearly, consider that if a realist philosopher of mathematics is asked what, in particular, arithmetical reality consists of, he might reply that arithmetical reality is made of abstract objects we call 'natural numbers' and that arithmetic is the science of such objects.[4]

However, at this point, another discussant of realist inclinations could observe that since the natural numbers can be determined only up to isomorphism, arithmetic should not be seen as a science of abstract objects (which objects?), but, rather, as a science describing that complex abstract entity—comprising abstract objects and relations defined on them—which has come under the name of 'arithmetical structure;' and that, if we wanted to picture such a structure, we could imagine it as a web in which the nodes are natural numbers, and the links between them are directed edges representing relations holding between natural numbers.

1.4 On Mathematical Structure

The controversy between realists about objects and realists about structures in the philosophy of mathematics is relatively recent. Its beginnings may be traced back to the very end of the nineteenth century. When, in 1899, Hilbert published *The Foundations of Geometry*,[5] Frege observed that in his book Hilbert had given no definitions of point, line, plane and of the betweenness relation.[6] Hilbert's reply to Frege, besides abolishing the traditional strict distinction between definitions and

[4] See on this Frege (1884), § 61.

[5] See Hilbert (1899a).

[6] Frege (1899).

axioms by means of the introduction of implicit definitions,[7] was given in the purest structuralist style. And the gist of it was that he did not care what points, lines, planes and betweenness were in so far as these objects (and relation) obeyed the axioms of Euclidean geometry he had laid down in his book.[8]

Despite the suggestiveness of Hilbert's view of Euclidean geometry, and the structuralist development of twentieth century mathematics for how this is, for instance, manifested in the work of the Bourbakists, an important open question which is still at the heart of any structuralist philosophy of mathematics is providing a satisfactory definition of mathematical structure. What we said on this topic in Sect. 1.3 was far too impressionistic (the web ...) and vague to be of much use.

In any case, whatever the correct definition of mathematical structure might be, a mathematical structure has to be something that bears a strong resemblance to the Aristotelian concept of form as what individuates and determines a given (mathematical) entity above and beyond the objects of which this is made. 'Pattern,' we think, would be a good word for it.

But, is the absence of a unique, shared definition of mathematical structure a problem for a structuralist philosophy of mathematics? We think not, in so far as we have very general examples of structures, such as the set-theoretical and category-theoretical structures, which are virtually applicable to the whole of mathematics. Why should there be only one kind of general mathematical *ur*structure from which all other mathematical structures must originate?

1.5 Logics and Metaphysics

For someone used to thinking of logic as the science of deductive thought, it is not easy to come to terms with the existence of various different, and mutually incompatible, logical systems and, in particular, with what the laws of deductive thought might have to do with the existence/non-existence of God, time, mathematical reality, and various other important objects of philosophical investigation falling under the concept of metaphysics, broadly construed. The aim of this section is to go some way towards an explanation.

The connection between logic and metaphysics goes back to the beginnings of Western philosophy. A particularly interesting example of such a connection is to be found in the philosophy of Heraclitus of Ephesus.

[7] An implicit definition of a mathematical entity G is a definition of this entity given by a set of axioms which characterise it. Well known examples of important implicit definitions in contemporary mathematics are those of group, ring, field, vector space, topological space, etc.

[8] See Hilbert (1899b).

If the concept of flux/change is an essential component of our thought about the world, we might be tempted to say, with Heraclitus, that:[9]

> We step and do not step in to the same rivers; we are and are not.

and, consequently, we might be drawn to accepting the Law of Contradiction $P \wedge \neg P$ (LC). As a matter of fact, if we study the history of Western philosophy, we realise that, apart from Heraclitus, several philosophers accepted LC, philosophers among whom we find Hegel, the dialectical materialists, and the so-called 'dialetheists,' that is, those logicians who believe in the existence of true contradictions (P and $\neg P$ are both true).

A traditional way of attacking metaphysical systems which compel the acceptance of LC was the discovery, due probably to Duns Scoto (1265–1308), of the so-called 'Principle of Explosion' of classical logic: *ex contradictione quodlibet*. This property, which classical logic shares with other logics, e.g., intuitionistic logic, has the effect of trivialising any formal system of classical logic $\mathfrak{F}$ within which we can prove a contradiction. It was only in the second half of the last century, with the advent of the so-called 'paraconsistent logic,' that a way of 'taming' the Principle of Explosion was found (in a non-classical logic). This, of course, ended up giving logical respectability to metaphysical systems like that of Heraclitus.

Another famous example of the interaction between logic and metaphysics comes from the Middle Ages. And it has to do with Anselm of Aosta's attempt to prove the existence of God[10]—one of the problems of traditional metaphysics[11]—by means of the so-called 'ontological argument.'

To put it succinctly, in the ontological argument Anselm argues from the *existence in the understanding* (possibility) of something than which nothing greater can be thought to the *existence in reality* (actuality) of such a being. Although the ontological argument appears to be in clear contrast with the celebrated logical principle *a posse ad esse non valet consequentia*, it has been an object of controversy from the time of its first publication to Kurt Gödel's repeated attempts to recast it in the shape of a formal proof of modal logic.[12]

[9] Diels and Kranz (1969), vol. 1, Chapter 22 Eraclito, Fragment 49a, p. 207.

[10] See on this Anselm (2007).

[11] See on this Kant (1787), I Transcendental doctrine of the elements, Second Division. Transcendental Dialectic, Book I, § 3, footnote a, p. 325:

> Metaphysics has as the proper object of its inquiries three ideas only: *God*, *freedom*, and *immortality* [...]

[12] See on this Gödel (2008).

More recent examples of the interaction of logic and metaphysics are to be found in the work of Kant[13] (classical logic); Arthur Prior[14] (tense logic); in the debate concerning whether or not one should be realist about possible worlds[15] (modal logic); and in Michael Dummett's way of formulating the realism/anti-realism debate (classical vs. intuitionistic logic).

In the course of the history of Western philosophy legions of philosophers of mathematics have been quarrelling over the existence of mathematical reality—Plato, Aristotle, Frege, Russell, Hilbert, Brouwer, Gödel, Quine, and countless others, all grappled with this issue at some point or another—but, it is in Dummett's peculiar way of recasting the traditional realism/anti-realism dispute about mathematics that we see one of the clearest examples of the interaction between various types of logic and metaphysics.

According to Dummett:

> [i]t is difficult to avoid noticing that a common characteristic of realist doctrines is an insistence on the principle of bivalence—that every proposition, of the kind under dispute,[16] is determinately either true or false. Because, for the realist, statements about physical reality do not owe their truth-value to our observing that they hold, nor mathematical statements their truth-value to our proving or disproving them, but in both cases the statements' truth-value is owed to a reality that exists independently of our knowledge of it, these statements are true or false according as they agree or not with that reality [...] Those who first clearly grasped that rejecting realism entailed rejecting classical logic were the intuitionists, constructivist mathematicians of the school of Brouwer.[17]

In what Dummett asserts in the quotation above not only do we see how different views concerning the existence of mathematical reality entail the acceptance of different types of logic, but we also come to realise that, perhaps, there is a way of translating the traditional metaphysical debate between realists and anti-realists in the philosophy of mathematics into a 'logical' debate on the nature of mathematical truth: agreement with reality or provability?[18]

One of the important consequences of such a translation would be that, since in a debate about the nature of mathematical truth language looms large—after all,

[13] Consider Kant's famous remark:

> '*Being*' is obviously not a real predicate [...] Logically, it is merely the copula of a judgment. (Kant (1787), I Transcendental doctine of elements, Second Part, Second Division, Book II, Chapter III, § 4, p. 504.)

For a modern analysis of the consequences of Kant's view of 'being' on the ontological argument and on Descartes' *cogito* argument see Carnap (1932).

[14] See, for instance, Prior (1957).

[15] Such a debate sees modal realists such as David Lewis (see Lewis 1986) opposing modal anti-realists like Saul Kripke (see Kripke 1980).

[16] In a Dummettian realism/anti-realism debate the propositions under dispute are those and only those for which both the realist and anti-realist admit that they do not have criteria of decision. See on this Oliveri (1994).

[17] Dummett (1991), Introduction, p. 9. On this see also Dummett (1973).

[18] See on this Dummett (1959) and Dales and Oliveri (1998).

truth can always be thought of as the truth of a proposition—it seems, in principle, possible to develop a theory of meaning[19] for the language of mathematics which might have a chance of adjudicating the dispute.

1.6 Logics and Ontology

A classic example illustrating the importance of logic for the debate on the nature of the entities studied by mathematical theories is the dispute about whether mathematics is a science of objects or structures.

As we shall see in this section, the discussion concerning what kind of (classical) logic—first- or second-order—one should adopt in developing formal systems of arithmetic might prove of some importance for the objects vs. structures controversy in the philosophy of mathematics.

Let T be a consistent formal system of arithmetic. As we have seen in Sect. 1.3, we can consider T as a science of objects, *à la* Frege, or as a science of structures, where by 'structure' here we mean a model M of T. If T is first-order, say T is Peano Arithmetic (**PA**), it will certainly have the standard (or natural) model in whose domain we have all the natural numbers, etc. But it will also have non-standard (or 'Frankenstein') models which are not isomorphic to the natural model of T.[20]

In this situation it comes easy to: (1) dismiss the view that T is a science of structure, because the criterion of identity for structures is structure isomorphism, and there are models of T which are not isomorphic to one another; (2) regard the Frankenstein models of T as interesting/annoying curiosities; and (3) cling on to the idea that arithmetic is *really* the science of those dear objects we have all become familiar with from the time of our 'gingerbread or pebble arithmetic'[21] at elementary school, that is, the objects we refer to by means of the numerals $0, 1, \ldots, n, n+1, \ldots$

However, if T is a consistent formal system of (full) second-order arithmetic then, from Dedekind's Categoricity Theorem, we know that any two models of T are isomorphic to one another. What this means is that there is *essentially* one structure satisfying the axioms of T, and we highlight this remarkable fact by saying that T is *categorical*.[22]

Clearly, the categoricity of T plays right into the hands of the structuralist philosopher of mathematics. For, if T is categorical, there is only one structure characterised by T and, consequently, the elements of the domains of the various models of T can be regarded as mathematically irrelevant. Now the structuralist philosopher of mathematics can smugly say, with Hilbert, that it does not matter

[19] See on this Dummett (1991), Introduction and Chapter 15.

[20] See on this Boolos and Jeffrey (1991), Chapter 17.

[21] Frege (1884), Introduction, p. VII.

[22] See on this Boolos and Jeffrey (1991), Chapter 18.

what kind of things the terms '0,' 'natural number,' and 'successor' refer to in so far as these entities satisfy the (full) second-order Peano axioms.

As the reader can easily see, arguing in favour/against the adoption of first-order/second-order logic in formal systems of arithmetic has profound consequences on the (ontological) objects vs. structure dispute about arithmetic.[23]

1.7 The Book

Objects, Structures and Logics is an edited collection of articles based on some of the talks given at the III International Conference of the Italian Network for the Philosophy of Mathematics (FilMat). The conference took place in the town of Mussomeli (Sicily) at the end of May 2018, and was a success both scientifically and socially.

The papers appearing in this volume have been arranged into three main sections. Although the sectioning adopted will inevitably look somewhat artificial, we believe that it provides the reader with a reliable indication of the book's main topics.

The first section, 'Mathematical Objects', is arguably the most metaphysical in character, featuring contributions which address the nature and definability of mathematical objects, and, among others, the long-standing questions of what mathematical objects there are, on what grounds some objects may be seen as having greater foundational relevance than others, and neo-Fregeanism.

The second section, 'Structures and Structuralisms' ties in with our discussion of Structuralism (Sect. 1.4), and of its philosophical ramifications. The papers included here assess the strength of several structuralist proposals, and examine novel approaches to well-established and more recent questions.

The third section, 'Logics and Proofs', comprises work focussing on century-old issues in the philosophy of mathematics, e.g., the nature and value of mathematical proofs (also in non-classical contexts), mathematical theories of truth, and explores the recently emerged proof-theoretic notion of *grounding*.

1.7.1 Part I: Mathematical Objects

In the first paper of the collection **Luca Zanetti** challenges the standard definition of mathematical Platonism according to which mathematical entities exist and are mind, and language, independent. Zanetti argues that the standard definition fails to distinguish Platonism from several varieties of Aristotelianism in the philosophy of mathematics; and that Platonism ought to be characterized rather in terms of meta-

[23] This is, by no means, a situation which concerns only arithmetic. Similar considerations are applicable to formal systems for the real numbers and other mathematical theories.

physical grounding, as the idea that there are fundamental mathematical entities. By contrast, says Zanetti, Aristotelianism upholds the existence of mathematical objects, even though these are not considered to be metaphysically fundamental.

Platonism and Aristotelianism share the view that mathematical objects do exist. Nominalism, by contrast, is the view according to which mathematical objects do not exist. **Eric Snyder**, **Richard Samuels** and **Stewart Shapiro** criticize nominalism by elaborating on the so-called 'Frege's Other Puzzle'. The Puzzle can be roughly stated as follows: how can an expression, such as 'four', serve seemingly different semantic functions in equivalent statements, such as 'Jupiter has four moons' and 'the number of Jupiter's moons is four'? The authors present the traditional responses to the Puzzle and elaborate on Hofweber's adjectival strategy. According to Hofweber, a statement such as 'Jupiter has four moons' is true in virtue of non-referential determiners. This claim, if plausible, would lead to the conclusion that numbers do not refer to mathematical objects. Snyder, Samuels and Shapiro examine two components of Hofweber's strategy (the syntactic and the semantic components) and argue that Hofweber's solution of Frege's Other Puzzle does not survive strict empirical scrutiny.

In the third paper of this section, **Ingo Blechschmidt** discusses peculiar mathematical objects known as *toposes* (as he says, the 'toposophic' landscape), of which he provides an exhaustive overview. Toposes first emerged in the work of Alexander Grothendieck in the 1950–1960s. Subsequently, they became the subject of independent investigation in the context of *category theory*. Indeed, as Blechschmidt shows, a topos shares many categorical properties with the *category of all sets*, that is, the category whose domain consists of all sets, and all maps among them. One further distinctive feature of toposes is that they standardly obey *intuitionistic*, rather than *classical*, logic; therefore, they may be more suitable than alternative foundational frameworks (such as *set theory*) to deal with a variety of mathematical universes arising not only from alternative choices of axioms, but also from different conceptions of logic. Among other things, Blechschmidt's paper examines the relevance of toposes to the debate on several outstanding issues in the contemporary philosophy of mathematics (some of which have been reviewed in Sect. 1.1): the extent and strength of non-classical logics, the realism/anti-realism dispute, the nature of *mathematical objects*, the relationship between toposes and mathematical (set-theoretic) universes (the set-theoretic multiverse), the problem of how mathematical structures should be characterised, and, finally, the viability of conceptions of non-standard analysis (such as *synthetic differential geometry*) alternative to Abraham Robinson's original system of hyperreals.

The neo-Fregean programme has recently emerged as an attempt to revive Frege's logicism. In his paper, **Daniel Waxman** takes into account the issue of the indispensability of 'implicit definitions' within such programme. As Waxman explains in detail, neo-Fregeans have come to view the logicist programme as hinged upon the definition of mathematical terms through implicit definitions. In turn, implicit definitions are commonly thought to be based on abstraction principles. The quintessential and, one should add, most prominent abstraction principle targeted by neo-Fregeans is Hume's Principle. Now, Waxman acknowledges that implicit

definitions, and abstraction principles, fulfil many tasks which are relevant to the neo-Fregean project. However, he explores a different strategy, the Hilbertian Strategy, which views the definition of mathematical terms as given by the axioms of mathematical theories themselves. Then, he proceeds to produce a full-fledged defence of this strategy against potential objections raised by such neo-Fregeans as Hale and Wright, and concludes that 'neo-Hilbertianism' might be a serious contender of abstraction principles.

1.7.2 Part II: Structures and Structuralisms

In the first contribution of this section **Neil Barton** examines the significance and strength of the structuralist viewpoint for contemporary set theory. The author first reviews classifications of structures into *general* (or *algebraic*, those which admit of non-isomorphic exemplars) and *particular* (or *non-algebraic*, those which do not admit of non-isomorphic exemplars), and then deals with the issue of how one comes to isolate and characterise specific mathematical structures. The author contrasts Georg Kreisel's view, that the identification of structures is carried out through a process of 'informal rigour', which culminates in the adoption of a categorical (second-order) set of axioms for the structure under consideration, with Michael Resnik's idea, which considers the characterisation of structures as being open to the use of different logical resources. In order to tackle the issue in more detail, Barton focuses on a paradigmatic case study, that of set theory, and argues that, while Kreisel's suggestion might be partly correct, in the sense that set-theoretic reference might be dictated by categoricity concerns as expressed by the idea of informal rigour, the logic underlying set-theoretic thought, which Barton suggests may be weaker than second-order logic, would still allow one to keep the algebraic interpretation of set theory alive. Thus, Barton concludes that our level of informal rigour within set theory is not high enough to decide whether, for instance, the Continuum Hypothesis is true or false.

In the subsequent paper **Silvia Bianchi** contends that Shapiro's structuralism can be interpreted in terms of grounding so as to avoid any eliminative view of mathematical objects. She advocates weak mathematical structuralism, that is, the idea that mathematical objects are metaphysically grounded in the structure they belong to. Bianchi starts by drawing a comparison between the philosophy of science and mathematics, and ends by defending a thin concept of mathematical objects. Bianchi attempts to vindicate the priority of structures over objects by overcoming the main objections to *ante rem* structuralism.

In his intriguing, if unconventional, contribution **Jean-Pierre Marquis** tackles the issue of whether a notion of 'mathematical style' can be successfully articulated. To this end, he examines a fascinating case study, that of Bourbaki's *Éléments*, a milestone in the history of the foundations of mathematics which sparked controversy since its appearance in the 1940s. Before plunging into the examination of his case study, Marquis provides a brief overview of alternative methods and ways

of doing mathematics which would exemplify the notion of style. He thus proceeds to outline the idiosyncratic elements of Bourbaki's conception of mathematical work, something which leads him to isolate what one may view as *bona fide* features of a 'Bourbakist style'. As Marquis clarifies, the latter would consist in a 'structuralist style', that is, a peculiar way of doing mathematical work based on: (i) viewing structures as fundamental objects of investigation, (ii) selecting the right axioms for them, and (iii) committing oneself to absolute logical rigour in proofs. Marquis' examination of the topic includes, among other things, useful information about the history and development of the still ongoing Bourbakist undertaking.

While Blechschmidt's paper provides a comprehensive account of topos theory, by exploring a broad range of philosophical scenarios and ramifications, **Olivia Caramello**'s contribution to this volume focuses on a specific class of mathematical *structures*, that is, Grothendieck toposes. In particular, Caramello defends the idea that the latter may be viewed as 'unifying bridges' in mathematics. First, Caramello explains that the kind of unification she wishes to investigate does not just reduce to mere generalisation through 'dilution' of the pre-existing mathematical objects, but is, rather, a way of connecting objects in a dynamic way (that is, a way which allows one to transfer bits of knowledge from a given class of objects to another one), by pinning down special abstract properties which underlie the objects under consideration, an approach which, among other things, clearly expresses *structuralist* concerns. In more practical terms, this 'unification' consists in identifying high-level invariants, of which the objects connected may, in turn, be seen as low-level expressions, or forms, a process which Caramello describes as *morphogenesis*. Afterwards, she proceeds to exhibit concretely how Grothendieck toposes, arising from a generalisation of other concepts of algebraic geometry, a discipline having noticeable unifying virtues itself, help 'make bridges' among different mathematical objects, and reviews other relevant mathematical phenomena which make sense of this perspective.

1.7.3 Part III: Logics and Proofs

The first article of this section, authored by **Davide Catta** and **Antonio Piccolomini D'Aragona**, makes an original comparison between Jean-Yves Girard's 'ludics' and Dag Prawitz's 'theory of grounds', emphasising some shared philosophical ideas about proofs and deduction. Catta and Piccolomini then proceed to outline a formal translation of the implicational fragment of intuitionistic logic. This translation, they suggest, may lead to a dialogical reading of Prawitz's theory of grounds. The translation, although sketchy, is an important starting point to frame intuitionistic logic within Girard's ludics.

Intuitionism is a form of constructivism, which in general is concerned with constructive mathematical objects and reasoning. Constructive mathematics is a form of mathematics that uses intuitionistic rather than classical logic. Along with constructivism, predicativism regards as suspect talk of definitions that attempt

to define mathematical entities circularly. In her contribution **Laura Crosilla** elaborates on the notion of predicativity in the foundations of mathematics. The standard approach to predicativism is based on formal theories that reject generalised inductive definitions. This approach originates from Russell and Weyl, in particular, from the idea that sets of natural numbers need to be defined predicatively to avoid any violation of the Vicious–Circle Principle. It is, perhaps, overlooked that inductive definitions could be predicative in a sense that differs from the (standard) proof–theoretic analysis of predicativity. Crosilla offers three possible strategies that are available to the constructivist to defend the predicativity of inductive definitions.

The notion of truth plays a central rôle in the relation between logic and the philosophy of mathematics. Indeed, a strong reason to commit ourselves to the existence of mathematical entities is provided by a certain view of mathematical truth. And, in turn, mathematical truth has been traditionally correlated to proof. **Andrea Cantini** examines the impact that the truth predicate has on the philosophy of mathematics by presenting axiomatic theories of truth leading to a brand-new analysis of predicativity and truth predicates.

Enrico Moriconi's paper addresses Imre Lakatos' notion of proof. As is well known, Lakatos' view was that mathematical proofs do not just reduce to formal procedures. In fact, formalisation is just one of the many aspects involved. Other aspects of 'proof-making' also include the creation of new concepts (sometimes through the 'stretching' of existing ones); moreover, proof-making is compared by Lakatos to the fallibilist process consisting in taking into account alternative hypotheses and selecting that which best accommodates the available data, a view which makes Lakatos' doctrines resemble Popper's own falsificationist conception in the philosophy of science. Moriconi explores both Lakatos' notion of 'concept-stretching', by illustrating the Hungarian philosopher's thesis that semantic and definitional aspects of proof-making may not be separated from the purely formal, proof-theoretic ones, as well as the controversial relationship between Popper's and Lakatos' doctrines, by highlighting their multiple convergences and parallel evolution.

Finally, in the last contribution to this section **Samuele Maschio** addresses 'existence properties' in the context of constructive mathematics. Such properties are easily definable in a more powerful 'ambient' theory, that is, category theory, and Maschio's goal is precisely that of examining category-theoretic versions of the properties in question. To this end Maschio first describes the semantics underlying constructive mathematics known as the BHK interpretation of the logical constants (after Brouwer, Heyting and Kolmogorov), and then proceeds to reformulate existence properties in terms of metamathematical properties. This can, practically, be done by means of a translation of such properties into definable classes, and then taking into account the category of all such definable classes. The effectiveness of the translation also exhibits the versatility of category theory, which, like topos theory, allows one to take into account a broad range of alternative logics, axioms and mathematical universes.

Acknowledgments The successful outcome of FilMat's III International Conference was the result of a collective effort involving a considerable number of individuals and institutions. In particular, we would like to thank: the FilMat Network, for suggesting to one of the editors of this volume that the conference be held in Sicily; the town council of Mussomeli, the University of Palermo, the Regione Siciliana, the Accademia Nazionale di Scienze Lettere ed Arti di Palermo, the Associazione Italiana di Logica e sue Applicazioni, the Società Italiana di Logica e Filosofia della Scienza, and the Società Italiana di Filosofia Analitica, for sponsoring the event; the people from Mussomeli who made sure that the day-to-day logistics worked smoothly; all the speakers, for their interesting and engaging talks; Francesca Boccuni, Jessica Carter, Julian Cole, Alain Connes, Thierry Coquand, Leo Corry, Giovanna Corsi, Marcello D'Agostino, Ciro De Florio, Ali Enayat, Ivan Fesenko, Salvatore Florio, Donald Gillies, Geoffrey Hellman, Thomas Hofweber, Laurent Lafforgue, Elaine Landry, Øystein Linnebo, Paolo Mancosu, Tony Martin, Colin McLarty, Marco Panza, Francesco Paoli, Frédéric Patras, Mario Piazza, Paolo Pistone, Matteo Plebani, Erich Reck, Helmut Schwichtenberg, Andrea Sereni, Luca Tranchini, Jaap van Oosten, Crispin Wright, for contributing, in various capacities, to the scientific level of the conference and of this volume.

A vital rôle in the preparation and organisation of the conference was that of Prof. Roberto Prisco from Mussomeli. His contagious enthusiasm for the organisation of important cultural events, and ability in convincing others, have proved invaluable both for the choice of Mussomeli as the conference venue and for the involvement of the town council of Mussomeli as a sponsoring body of the meeting.

These acknowledgements would be seriously incomplete, if we did not express our gratitude to two anonymous referees for their helpful comments and to Christopher Wilby, Prasad Gurunadham and Svetlana Kleiner of Springer, our publisher, for having shown an interest in our publication project and much understanding of the difficulties we have come across in these strange, trying times.

References

Anselm of Aosta: 1077–78. 2007. *Proslogion*, ed. L. Pozzi, $VIII^{th}$ ed. Milano: BUR.

Benacerraf, P. 1965. What numbers could not be. In Benacerraf and Putnam (1985), Part II, 272–294.

Benacerraf, P., and H. Putnam, eds. 1985. *Philosophy of Mathematics, Selected Readings*, 2nd ed. Cambridge: Cambridge University Press.

Boolos, G.S., and R.C. Jeffrey. 1991. *Computability and Logic*, 3rd ed. Cambridge: Cambridge University Press.

Carnap, R. 1932. Il superamento della metafisica mediante l'analisi logica del linguaggio. In Pasquinelli (1969), 504–532.

Dales, H.G., and G. Oliveri. 1998. *Truth in Mathematics*. Oxford: Oxford University Press.

Diels, H., and W. Kranz. 1969. *I Presocratici*, vols. 1 and 2, ed. G. Giannantoni. Bari: Editori Laterza.

Dummett, M.A.E. 1959. Truth. In Dummett (1980), 1–24.

Dummett, M.A.E. 1973. The philosophical basis of intuitionistic logic. In Benacerraf and Putnam (1985), Part 1, 97–129.

Dummett, M.A.E. 1980. *Truth and Other Enigmas*, 2nd printing. Cambridge, MA: Harvard University Press.

Dummett, M.A.E. 1991. *The Logical Basis of Metaphysics*. London: Duckworth.

Dummett, M.A.E. 1994. Reply to Oliveri. In McGuinness and Oliveri (1994), 299–307.

Frege, G. 1884. *The Foundations of Arithmetic*, 2nd Revised Edition, English Translation by J.L. Austin, 1980. Evanston: Northwestern University Press.

Frege, G. 1899. Letter of Frege to Hilbert 27 December 1899. In Frege (1977), 458–462

Frege, G. 1977. *Logica e Aritmetica*, ed. C. Mangione. Torino: Boringhieri.

Gödel, K. 2008. *La prova matematica dell'esistenza di Dio*, ed. Gabriele Lolli e Piergiorgio Odifreddi. Torino: Bollati Boringhieri.
Hilbert, D. 1899a. *The Foundations of Geometry*, Translated by E.J. Townsend, Reprint Edition, 1950. La Salle: The Open Court Publishing Company.
Hilbert, D. 1899b. Letter of Hilbert to Frege 29 December 1899. In Frege (1977), 462–466.
Kant, I. 1787. *Critique of Pure Reason*, 2nd ed. Translated by Norman Kemp Smith, 1990. London: Macmillan.
Kripke, S.A. 1980. *Naming and Necessity*. Oxford: Basil Blackwell.
Lewis, D.K. 1986. *On the Plurality of Worlds*. Oxford: Oxford University Press.
Maddy, P. 2011. *Defending the Axioms*. Oxford: Oxford University Press.
McGuinness, B.F., and G. Oliveri, eds. 1994. *The Philosophy of Michael Dummett*, Synthese Library, vol. 239. Dordrecht/Boston/London: Kluwer Academic Publishers.
Meadows, T. 2013. What can a categoricity theorem tell us? *Review of Symbolic Logic* 6(3): 524–544.
Oliveri, G. 1994. Anti-realism and the philosophy of mathematics. In McGuinness and Oliveri (1994), 379–402.
Oliveri, G. 1997. Mathematics. A science of patterns? *Synthese* 112(3): 379–402.
Pasquinelli, A., ed. 1969. *Il Neoempirismo*. Torino: U.T.E.T.
Prior, A. 1957. *Time and Modality*. Oxford: Oxford University Press.
Shapiro, S. 1997. *Philosophy of Mathematics. Structure and Ontology*. Oxford: Oxford University Press.
Strawson, P.F. 1974. *Individuals*. London: Methuen.
van Heijenoort, J., ed. 1967. *From Frege to Gödel*. Cambridge, MA: Harvard University Press.
Zermelo, E. 1908. Investigations in the foundations of set theory I. In van Heijenoort (1967), 199–215.

Part I
Mathematical Objects

Chapter 2
Aristotle's Problem

Luca Zanetti

Abstract Platonism is traditionally defined as the view that there are abstract mathematical objects, and that those objects are independent of human beings and their thoughts, language, and practices. This paper has two goals. First, to show that this definition fails to distinguish platonism from various forms of *aristotelianism* in the philosophy of mathematics, according to which mathematical objects depend for their existence and properties on non-mathematical ones. Second, to argue that platonism is best defined in terms of *metaphysical fundamentality*, as the view that there are fundamental mathematical entities. I finally distinguish between different varieties of mathematical aristotelianism.

Keywords Mathematical platonism · Aristotelianism in mathematics · Fundamentality · Metaphysical dependence · Ontological commitment

2.1 Introduction

You know one from the company it keeps. For this reason, we will start our investigation of *aristotelianism* in mathematics from platonism. As emphasized by Linnebo (2018, 189), Mathematical platonism ('platonism' for short) is traditionally characterized by two claims:

> [EXST]
> There exist abstract mathematical objects.

I wish to thank Øystein Linnebo, Andrea Sereni, the Editors of this volume, and two anonymous referees for their comments, which have greatly improved this work.

L. Zanetti (✉)
NEtS Center, Department of Humanities and Life Sciences, Scuola Universitaria Superiore IUSS Pavia, Pavia, Italy
e-mail: luca.zanetti@iusspavia.it

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_2

[IND]
Mathematical objects, if any, exist and have their properties independently of intelligent agents and their language, thought, and practices.

Let's consider those two claims in some details.

EXST consists of two sub-claims: (1) mathematical objects exist, and (2) those objects are abstract entities.

As regards (1), the claim that mathematical objects exist can be usefully explicated in terms of the *ontological commitments* of platonism. According to the well-known criterion due to Quine, the ontological commitment of a theory consists in what must lie in the range of its (first-order) quantifiers in order for the statements of that theory to be true. Quine's criterion can be made tolerably clear by means of the standard notions of interpretation, satisfaction, and logical entailment.

As usual, let an *interpretation* of a first-order language be a pair $M = \langle d, I \rangle$, where d is an non-empty set, i.e. the domain, and I is a function which assigns a member of d to each first-order constant of the language, a subset of d to each one-place predicate letter, and a set of ordered n-tuples of members of d to each n-place predicate letter; and let an assignment on an interpretation M be a function which assigns a member of d of M to every first-order variable. For each singular expression t, if t is a constant, let the denotation of t in M under the assignment s ('$D_{M,s}(t)$') be $I(t)$, and, if t is a variable, let $D_{M,s}(t)$ be $s(t)$. If t_1 and t_2 are terms, then M *satisfies* the formula '$t_1 = t_2$' ('$M, s \models t_1 = t_2$') if and only if ('iff') $DM, s(t_1)$ is the same as $DM, s(t_2)$. If R^n is an n-place predicate letter and $t_1, \ldots, t_n$ are singular expressions, then $M, s \models$ '$R^n(t_1, \ldots, t_n)$ iff the n-tuple $[DM, s(t_1), \ldots, DM, s(t_n)] \in I(R^n)$. Satisfaction for complex sentences is defined inductively as usual. Let ϕ and ψ be schemas for sentences of the language. $M, s \models \neg(\phi)$ iff it is not the case that $M, s \models \phi$; $M, s \models \phi \wedge \psi$ iff both $M, s \models \phi$ and $M, s \models \psi$; $M, s \models \phi \vee \psi$ iff either $M, s \models \phi$ or $M, s \models \psi$; $M, s \models \phi \rightarrow \psi$ if and only if it is not the case that $M, s \models \phi$, or $M, s \models \psi$. Finally, $M, s \models \forall u\ (\phi)$ if and only if, for every assignment s' that agrees with s except possibly at the variable u, $M, s' \models \phi$, and $M, s \models \exists u\ (\psi)$ iff, for some assignment s' that agrees with s except possibly at the variable u, $M, s' \models \phi$. We say that ϕ is logically entailed by a set of sentence Γ if, for any interpretation M of the language, if $M \models \psi$, for each $\psi \in \Gamma$, then $M \models \phi$.

Given these notions, Quine's notion of ontological commitment can be glossed as follows: a theory T is ontologically committed to K's if and only if it logically entails '$\exists x\ (K(x))$', i.e. if and only if for any interpretation which satisfies the sentences of T, there is some entity in the domain of interpretation which belongs to the extension of K. Given this clarification, (1) can be formalized as:

$$\exists x\ (MAT(x)),$$

where '$MAT(x)$' is a predicate which is true of all and only mathematical objects. Given our clarifications above, platonism is the view that is ontologically committed to mathematical objects.

As regards (2), it is generally assumed that mathematical objects are abstract rather than *concrete* ones. Lewis (1986) famously distinguishes four ways in which the distinction between abstract and concrete entities can be made, which he labels the *Way of Negation* (WoN), the *Way of Example* (WoE), the *Way of Conflation* (WoC), and the *Way of Abstraction* (WoA). On the WoN, abstract entities are characterized by the properties that they typically lack; in particular, abstract objects lack spatio-temporal location (abstract objects are nowhere) and causal powers (nothing is caused by abstract entities, and they are not caused by anything). On the WoE, abstract entities are those that are similar, in some sense to be specified, to paradigmatic cases of abstractness, e.g. sets. On the WoC, the distinction between concrete and abstract entities is explained by recourse to more familiar or better understood distinctions (e.g. between particulars and universals). On the WoA, abstract entities are the result of some mental process, i.e. abstraction, which is usually taken to consist in forming a general concept by omitting some of the features which distinguish particular objects from one another.

Lowe (1995) makes a similar distinction between three conceptions of abstractness. In particular, *abstract*$_1$ *objects* are "nonspatiotemporal in nature", while concrete objects are "thought of as existing in space and time" (p. 513); *abstract*$_2$ *objects* are "logically incapable of enjoying a 'separate' existence ... even though they might be separated 'in thought'" (p. 514); finally, *abstract*$_3$*objects* are "introduced by way of *abstraction from concepts*, according to Fregean abstraction principles" (pp. 514–4).

Lowe's distinction partially overlaps with Lewis's one. Lewis's WoA and Lowe's *abstractness*$_2$ correspond to a *psychological* conception of abstraction, according to which abstract entities result from some kind of mental process. By contrast, Lowe's *abstractness*$_3$ corresponds to a *logical* conception of abstraction, which consists in an assignment of objects to (possibly non-abstract) items on the basis of a given equivalence relation over those items. More precisely, a (Fregean[1]) abstraction principle is a universally quantified biconditional of the form '$\forall\alpha\forall\beta\ (\Sigma(\alpha) = \Sigma(\beta) \leftrightarrow \alpha \sim \beta)$', where α and β are variables of the same type (e.g., first-order or second-order), 'Σ' is a term-forming operator that denotes a function from entities of the type of α and β to objects, and $\sim$ stands for an equivalence relation over entities of the given type.

Frege (1953) gave two famous examples of abstraction. The first one concerns *directions*; the principle states that for any two lines a and b, those lines have the same direction if and only if they are *parallel*. Frege's second example is *Hume's Principle* (HP), which states that for any two concepts F and G, the *cardinal number* of F ('$\#(F)$') is the same as the cardinal number of G ('$\#(G)$') just in case F and G are *equinumerous*, i.e. if and only if the F's and the G's can be put into one-to-one correspondence ('$F \approx G$'), where '$F \approx G$' abbreviates the (purely second-order) statement that there is a relation R such that every object falling under F is R-

[1] Cf. Frege (1953, §§ 64-7); see Mancosu (2016) for an overview of the use and significance of abstraction principles before and after Frege.

related to a unique object falling under G, and every object falling under G is such that there is a unique object falling under F which is R-related to it.

In this paper we will take 'abstract object' to mean 'object by abstraction', i.e. objects that are assigned items of a given type by an abstraction function introduced by the relevant Fregean principle.[2] We will consider the second claim, IND, in the next section.

2.2 The Independence Thesis

The independence claim states that mathematical objects are independent of human beings and their thoughts, language, and practices.

This claim is informed by an analogy with ordinary physical objects; in the words of Linnebo (2009, Sect. 4.2), "[according to Platonists] just as electrons and planets exist independently of us, so do numbers and sets". More precisely, the analogy in question is meant to distinguish between mathematical objects, on the one side, and mind-dependent objects, like promises or institutions, on the other.

The independence claim is naturally glossed in modal terms as follows: had human thoughts, language and practice been different, or had there been no intelligent agent at all, there would still have been mathematical objects. Linnebo highlights that this claim informs our ordinary thinking about counterfactual scenarios:

> We often reason about scenarios that aren't actual. Were we to build a bridge across this canyon, say, how strong would it have to be to withstand the powerful gusts of wind? Sadly, the previous bridge collapsed. Would it have done so had the steel girders been twice as thick? This form of reasoning about counterfactual scenarios is indispensable both to our everyday deliberations and to science. The permissibility of such reasoning has an important consequence. Since the truths of pure mathematics can freely be appealed to throughout our counterfactual reasoning, it follows that these truths are counterfactually independent of us humans, and all other intelligent life for that matter. That is, had there been no intelligent life, these truths would still have remained the same.[3]

Let provisionally express IND as the counterfactual

[IND]
$\neg\exists x\ (MIND(x))\Box\!\!\rightarrow\exists x\ (MAT(x))$,

where, as before, 'MAT' is true of all and only mathematical objects, and '$MIND$' is true of all and only minds and mind-dependent entities (we also assume that $MIND$ and MAT are disjoint). IND states, therefore, that had there been no *minds*, there would still have been numbers.

[2] We will not consider, in particular, *in re* structuralism as a variety of mathematical aristotelianism; cf. Resnik (1997).

[3] Cf. Linnebo (2009, Sect. 4.1).

One might wonder whether this counterfactual conditional exhausts the intended analogy between physical objects and mathematical objects. If it did, then, as Linnebo remarks, anyone who accepts (mathematical) object realism, i.e. the view that there are abstract mathematical objects, should also accept IND. However, the independence claim is meant to express the more substantial idea that "mathematical objects are just as "real" as ordinary physical objects".

Moreover, one might also wonder whether the analogy with physical objects exhausts the Platonist notion of independence. Platonists assert not only that mathematical objects are mind- and language-independent, but also that they constitute a realm which is *separate* from the realm of both ordinary physical objects and mind-dependent entities. A plausible sense in which this claim is made is that mathematical objects entertain no causal or spatio-temporal relations with (our) world. At the same time, Platonists seem to put forward a more general metaphysical thesis, namely that mathematical objects do not owe their existence to non-mathematical ones; in particular, since mathematical objects are not ontologically dependent on neither intelligent agents nor ordinary physical objects, they would have existed, and been the same, even if there had been no intelligent agents or physical objects at all.

Finally, EXST and IND together fail to distinguish platonism and various forms of *aristotelianism* in the philosophy of mathematics.

In *Metaphyics* (XIII, 1076a, 35–37) Aristotle writes:

> If the objects of mathematics exist, they must exist either in sensible objects, as some say, or separate from sensible objects (and this also is said by some); or if they exist in neither of these ways, either they do not exist, or they exist only in some special sense. So that the subject of our discussion will be not whether they exist but how they exist.[4]

Later on, Aristotle asserts that "it is true to say without qualification that the objects of mathematics exist, and with the character ascribed to them by mathematicians". This leaves the mathematical realist with two options: either (A) numbers exist *"in"* physical things, or (B) they are *"separate"* from those things. Aristotle's option (B) corresponds to (traditional) platonism, which is in turn specified, as seen, in terms of EXST and IND. (A) is, by contrast, a form of *aristotelianism* in mathematics, according to which mathematical objects depend, for their existence and their properties, on non-mathematical entities.[5]

Here is a challenge to the standard definition of platonism. Indeed, Aristotle's options cannot be distinguished from each other on the basis of EXST and IND alone. On the one hand, the Platonist and the Aristotelian agree that mathematical objects exist. They might also agree, on the other hand, that mathematical objects

[4] Aristotle (1924, II, p.433).

[5] Aristotle's mathematical realism was in fact neither of type (A) nor of type (B); Aristotle held, by contrast, that mathematical entities are attributes of sensible objects, but those entities do not exist *in* those objects. My use of the term 'Aristotelianism' to indicate type-(B) mathematical realism traces back to Horsten and Leitgeb (2009, 217-8); cf. also Schwartzkopff (2011) and Donaldson (2017).

are mind- and language-independent.[6] However, they disagree on *how* those objects exist: according to the Platonist, mathematical objects are ontologically independent entities, while, according to the Aristotelian, those objects depend for their existence on physical ones.

2.3 Defining Platonism

The second goal of this paper is to argue that platonism is best defined in terms of *metaphysical fundamentality*.

Recall that, according to the Platonist, numbers exist independently of entities of other (i.e. non-mathematical) sort. It is therefore tempting to reformulate IND as follows:

> [SORT]
> For any non-mathematical sort K, mathematical objects exist and have their properties independently of whether K's exist.

The question is how the relevant notion of (in)dependence should be understood. It is natural to interpret this claim as before, by cashing it out in modal terms:

> [MODAL SORT]
> For any sort K of non-mathematical entities,
>
> $$\neg\exists x\ (K(x))\Box\!\!\rightarrow\exists x\ (MAT(x))$$

However, this modal gloss to SORT is insufficient, since it fails again to distinguish platonism from various forms of aristotelianism.

One such conception is Linnebo's *minimalism* in the philosophy of mathematics. Minimalism is the view that "mathematical objects are *thin* in the sense that very little is required for their existence" (2018, 3). More precisely, Linnebo claims that the truth of the right-hand side of an abstraction principle is *sufficient*, in a technical sense,[7] for the truth of its left-hand side.

Consider however the following instance of HP, whose right-hand side asserts that the concept *being non-self-identical*, $\ulcorner x \neq x \urcorner$, is equinumerous with itself:

$$\#(\ulcorner x \neq x \urcorner) = \#(\ulcorner x \neq x \urcorner) \leftrightarrow \ulcorner x \neq x \urcorner \approx \ulcorner x \neq x \urcorner \tag{0}$$

Following Linnebo's characterization, the right-hand side of (0) is sufficient for the truth of its left-hand side, and therefore for the number of not-self-identical

[6] Of course, the Aristotelian might alternatively claim that mathematical objects are mind-*dependent*; my point here is merely that a form of Aristotelianism which accepts both EXST and IND is a lively possibility, as witnessed by the examples below.

[7] Cf. Linnebo (2018, 140).

things, namely zero, to exist. However, it is a truth of (second-order) logic that $\ulcorner x \neq x \urcorner \approx \ulcorner x \neq x \urcorner$. Logical truths are typically conceived of as required nothing of the world for their truth: had there been no objects at all, and, in particular, had there been no non-mathematical objects, it would still have been the case that the concept *being non-self-identical* is equinumerous with itself, and, therefore, the number zero would have existed even if there had been no non-mathematical objects at all. Minimalism hence complies with MODAL SORT, even if Linnebo stresses that his own brand of lightweight platonism is less demanding than more traditional forms of mathematical realism, according to which the existence of mathematical objects imposes a significant burden on reality (cf. Linnebo (2018, xi).).

MODAL SORT must be strengthen in order to avoid this difficulty. The Aristotelian argues that mathematical objects are not among the most basic constituents of reality; she contends that numbers depend for their existence and their properties on entities of other kinds. The Platonist, by contrast, would argue that reality is, at least to some extent, *fundamentally mathematical*. Therefore, it might be reasonable to substitute IND, in any of its forms, with the following thesis:

> [FUND]
> There are fundamental mathematical objects.

The Platonist would assert FUND, while the Aristotelian would deny it. Note however that FUND might come in two versions. Let $FUND(x)$ be a predicate that is true of all and only the fundamental things.[8] The first and more radical view would be as follows:

> [STRONG FUND]
> $\forall x\ (MAT(x) \rightarrow FUND(x))$

For example, Shapiro (1997, 73-4) claims that the Platonist is committed, or might be committed, to the view that mathematical object, e.g. natural numbers, are independent of each other:

> [the Platonist] might attribute some sort of ontological independence to the individual natural numbers. Just as each beach ball is independent of every other beach ball, each natural number is independent of every other natural number. Just as a given red beach ball is independent of a blue one, the number 2 is independent of the number 6.

However, STRONG FUND might be too extreme. For example, the (traditional) Platonist might contend that the following two views are both correct:

(a) the empty set is ontological fundamental;
(b) each non-empty set ontologically depends on its members.

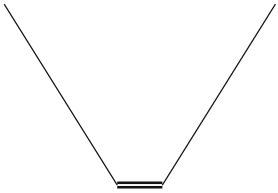
(a) and (b) entail that each (non-empty) pure set, that is, each set whose members are all sets, is an ontologically dependent entity. It seems therefore sensible to ascribe to the Platonist a more moderate view, that is,

[8] Nothing hinges here on whether the fundamentality is taken as a monadic or a as a relational property; cf. Wilson (2014, IV.i).

[WEAK FUND]
$\exists x\ (MAT(x) \wedge FUND(x))$

For example, the Platonist would argue that at least the empty set is fundamental, while the Aristotelian would claim that each set-theoretical object, including the empty set, depends for its existence on some non-mathematical entity or fact.[9]

WEAK FUND would moreover restore the analogy between physical and mathematical entities: the Platonist might indeed argue that both in the physical and in mathematical case, some entities are fundamental, while all the others depend on the fundamental base; the common structure that the Platonist would ascribe to both the physical realm and the mathematical realm is pictured below:



The Aristotelian would contend, by contrast, that while reality might contain physical entities at its fundamental level, mathematical objects should be placed at upper levels of the metaphysical ladder: the Aristotelian would argue, that is, that there are no mathematical objects *at the fundamental level*, or, more in general, that mathematical objects are not *as fundamental as* physical entities; the analogy between physical and mathematical objects would therefore break down.

If this is right, it also shows that the putative difference between the Platonist, which asserts WEAK FUND, and the Aristotelian, that denies it, can be very subtle, since they might disagree only on whether a single mathematical objects, e.g. the empty set, is fundamental. Provided that the Platonist accepts (b) – and, hence, that she denies STRONG FUND – the Platonist and the Aristotelian would only disagree on (a). However, diverging on (a) seems to be less a substantial difference between platonism and aristotelianism than what the traditional characterization of the former view might intuitively suggest.

[9] In order to make room for this latter view, fundamentality might be ascribed to *truths* rather than to objects. For a similar strategy, cf. Schwartzkopff (2011, 371-2).

2.4 Varieties of Aristotelianism

A reviewer asks whether there is any difference between aristotelianism as articulated here and more standard forms of mathematical naturalism.[10] My answer is that it depends on how both naturalism and aristotelianism are understood.

Naturalism in mathematics can be understood either as an ontological claim or as a methodological claim. *Ontological Naturalism* is the view that the objects of mathematics are natural. *Methodological Naturalism*, by contrast, is the view that the authoritative standards in the philosophy of mathematics are those internal to mathematics itself, those of natural sciences (e.g. physics), or both.

A form of methodological naturalism is usually advocated in order to defend mathematical realism. Paraphrasing Lewis (1991, 59), Johnathan Schaffer writes for example:

> Here, without further ado, is a proof of the existence of numbers:
>
> 1. There are prime numbers;
> 2. Therefore there are numbers.
>
> 1 is a mathematical truism. It commands *Moorean certainty*, as being more credible than any philosopher's argument to the contrary.[11]

Therefore, methodological naturalism seems to support *truth-value realism*, according to which mathematical theorems are in fact true. If one adds that mathematical statements like 1 and 2 should be interpreted with the logical form standardly assigned to them by mathematicians, then truth-value realism entails object realism, i.e. the view that there are mathematical objects.

However, platonism and aristotelianism supposedly agree on which mathematical objects exist. More precisely, both platonism and aristotelianism are plausibly committed to the existence of mathematical entities. So, the ontological commitments of platonism are just the same as the commitments of aristotelianism. Moreover, aristotelianism is also committed to the entities on which mathematical objects depend. Therefore, aristotelianism is *prima facie* even *less* ontological parsimonious than (traditional) platonism; even so aristotelianism would not be committed to *fundamental* mathematical objects.[12]

Whether *ontological* naturalism is entailed by aristotelianism will depend, on the other hand, on how the relevant notion of dependence between mathematical and non-mathematical objects is understood. We can indeed distinguish between (1) semantic, (2) truth-making, and (3) metaphysical varieties of aristotelianism.

[10] The reviewer's question was actually whether there is any difference between Aristotelianism and *empiricism*; I take empiricism in the sense of Lakatos (1976) to be a sub-species of naturalism, as I will explain below.

[11] Schaffer (2009, p. 357).

[12] Note, however, that aristotelianism may also be committed to *potential infinity*, whereas the Platonist isn't; if so, then the Platonist and the Aristotelian would not agree on EXST. I owe this comment to Gianluigi Oliveri.

(1) *Semantic aristotelianism* is the view that mathematical expressions should be treated as akin to free variables rather than proper names. According to Pettigrew (2008), the (semantic) Aristotelian claims, for example, that the terms formed by means of the operator 'the number of' are free variables (or parameters or arbitrary constants), rather than proper names formed by a term-forming operator flanked by a predicate; arithmetical statements would then express generalizations over each system of objects with the appropriate structure, rather than truths about a system of *sui generis* natural numbers.

Semantic aristotelianism might be interpreted as a form of ontological naturalism (modulo an assignment of values to free variables), provided that the domain of interpretation contains only natural entities, and it does not bring about any commitment to the existence of abstract mathematical objects unless one is also and independently commitment to their existence. For this reason, any version of semantic aristotelianism which is compatible with ontological naturalism denies EXST.

(2) *Truth-making aristotelianism* is the view that mathematical statements are made true by non-mathematical entities. According to Cameron (2008), the claim that the right-hand side and the left-hand side of HP (can be stipulated to) have the same truth-conditions is better interpreted as saying that the right-hand side has the same truth-makers as the left-hand side. This opens up two alternatives for the mathematical realist:

> if numbers are needed to make the left hand side true then are needed to make the right hand side true, and hence aren't a *new* ontological commitment; if numbers aren't needed to make the right hand side true then they aren't needed to make the left hand side true, and aren't a *commitment* (and so, *a fortiori*, they are not a new commitment)[13]

The Aristotelian accepts the second one of these options: what makes true the right-hand side of HP, i.e. that F and G are equinumerous, also makes true the left-hand side, i.e. 'the number of F is identical with the number of G'. The truth-making Aristotelian, therefore, is not committed to the existence of abstract mathematical objects, but only to non-mathematical truth-makers for mathematical truths, plus to truth-making relations themselves (even if those relation are 'thin' in the sense that can be explained by the semantic properties of mathematical language).

The question is whether truth-making aristotelianism can underwrite any sensible version of mathematical realism. Cameron claims indeed that (i) the sentence 'there are numbers' is literally true, and, at the same time, that (ii) in spite of (i), one's theory can be free of any commitment to the existence of numbers, since numbers are not needed to make the sentence 'there are numbers' true.

(3) *Metaphysical aristotelianism* is the view that mathematical truths metaphysically depend on (as opposed to: semantically depend on) non-mathematical

[13] cf. Cameron (2008, 12).

truths (Rosen, 2011, 2016). The metaphysical Aristotelian might claim, for example, that the right-hand side of an instance of HP, which states that two concepts are equinumerous, metaphysically grounds its left-hand side, which states that the numbers of those concepts are identical. Unlike semantic aristotelianism, metaphysical aristotelianism interprets mathematical expressions as proper terms and term-forming operators; unlike truth-making aristotelianism, moreover, metaphysical aristotelianism is committed to the existence of abstract mathematical objects. However, this last variety of aristotelianism is also burdened by a commitment to 'thick' metaphysical relations obtaining between mathematical and non-mathematical objects, which cannot be reduced to the semantic properties of mathematical language.

Which one of those three positions should be preferred will largely depend on considerations of ontological parsimony. The notion of (relative) ontological parsimony can be introduced in terms of the ontological commitments of two rival theories T_1 and T_2; T_1 is more ontologically parsimonious than T_2 if the ontological commitment of T_1, i.e. the set of entities to which T_1 is ontologically committed, is a proper subset of the ontological commitment of T_2. Many subscribe to the principle according to which entities must not be multiplied without necessity. Less idiomatically, the principle states that ontological parsimony is a virtue which should guide our choice between rival theories; more precisely, *(i)* if a theory T_1 is more ontological parsimonious than a theory T_2, T_2 should not be adopted unless its higher ontological costs are matched by other theoretical virtues, and *(ii)* other things being equal, we should prefer the most parsimonious theory, namely T_1.

As seen, only metaphysical aristotelianism is committed to the real existence of abstract mathematical objects, while semantic aristotelianism and truth-making aristotelianism are compatible with the view that there are no (*sui generis*) mathematical entities. Therefore, semantic and truth-making aristotelianism would give one an advantage with respect to ontological naturalism; in particular, both the semantic Aristotelian and the truth-making Aristotelian can in principle contend (modulo an appropriate articulation of their view) that everything that exists belongs to the natural world of causes and effects. By contrast, the metaphysical Aristotelian can only argue for a *moderate* naturalistic claim, according to which everything is either natural, or *metaphysically grounded* by the natural (Rosen, 2016, 279). Vice versa, if aristotelianism is defined (also) in terms of the existence claim, according to which there are abstract mathematical objects, then only metaphysical aristotelianism complies with this characterization, while semantic aristotelianism and truth-making aristotelianism don't.

Note, moreover, that even if aristotelianism is *prima facie* less ontological parsimonious than (traditional) platonism, since it is committed both to mathematical objects and to the non-mathematical entities on which those objects depend, aristotelianism posits no *fundamental* mathematical object. Indeed, ontological parsimony *at the fundamental level* is often the main reason to prefer aristotelianism over platonism. Schaffer argues, for example, that only fundamental entities should not be multiplied without necessity, and, more precisely, that the maximization

of the derivative entities that a theory can posit *vis-à-vis* the minimization of the fundamental entities that the theory must posit is itself a theoretical virtue which should guide our preference for theories which are more parsimonious (only) as far as fundamental entities are concerned (Schaffer, 2015, 647). Once Schaffer's maxim is adopted, it gives one a rationale for preferring aristotelianism over traditional platonism; as Schwartzkopff summarizes the point,

> in recent times it has been proposed to rather adhere to a principle one could call Ockham's fundamental razor: *entia fundamentalia non sunt multiplicanda praeter necessitatem*. So inclined philosophers find no difficulty in extending a welcoming hand to all kinds of things. The only caveat is that (i) such objects be non-fundamental or derivative, and (ii) that they ultimately derive from something such philosophers regard as unproblematic. One such sense of the fundamental *vs.* derivative distinction can be found in the Aristotelian tradition, a tradition in which substances are characterized as ontologically independent (fundamental) objects whereas the properties that inhere in them are said to be ontologically dependent (derivative) objects.[14]

It is worth noting, however, that while semantic and truth-making varieties of aristotelianism posits no fundamental mathematical objects (since they need not posit mathematical entities at all), metaphysical aristotelianism is committed both to relevant non-mathematical entities *and* to metaphysical relations between those entities and mathematical objects that are derivative on these latter. Unlike traditional platonism, however, aristotelianism does not posit fundamental mathematical objects.

Let's take stock. Semantic aristotelianism and metaphysical aristotelianism are both compatible with mathematical naturalism, i.e. the view that mathematical objects are natural; metaphysical naturalism, by contrast, is only compatible with a qualified form of naturalism, according to which mathematical objects are either natural, or grounded in the natural.

2.5 Conclusions

Recall that mathematical platonism is traditionally defined as the view that (1) there are abstract mathematical objects, and that (2) those objects are independent of human beings and their thoughts, language, and practices. I showed that (1) and (2) fail to distinguish platonism from various forms of aristotelianism in the philosophy of mathematics, namely the view that mathematical entities depend for their existence and their properties on non-mathematical ones. I then argued that platonism is best defined in terms of metaphysical fundamentality, as the view that there are (at least some) fundamental mathematical objects.

[14] Schwartzkopff (2011, 353).

References

Aristotle. 1924. *Aristotle's Metaphysics: A Revised Text with Introduction and Commentary by W. D. Ross*. Oxford: Clarendon Press.

Bliss, R., and G. Priest, eds. 2018. *Reality and Its Structure: Essays in Fundamentality*. Oxford: Oxford University Press.

Cameron, R.P. 2008. Truthmakers and ontological commitment: Or how to deal with complex objects and mathematical ontology without getting into trouble. *Philosophical Studies* 140(1): 1–18.

Clark, M.J., and D. Liggins. 2012. Recent work on grounding. *Analysis Reviews* 72(4): 812–823.

Correia, F., and B. Schnieder. 2012. Grounding: An opinionated introduction. In *Metaphysical Grounding: Understanding the Structure of Reality*, ed. F. Correia and B. Schnieder, p. 1. Camrbidge: Cambridge University Press.

De Florio, C., and L. Zanetti. 2020. On the schwartzkopff-rosen principle. *Philosophia* 48(1): 405–419.

Donaldson, T. 2017. The (metaphysical) foundations of arithmetic? *Noûs* 51(4): 775–801.

Fine, K. 2012. Guide to ground. In *Metaphysical Grounding*, 37–80. Cambridge: Cambridge University Press.

Frege, G. 1953. *The Foundations of Arithmetic*. Evanston: Northwestern University Press.

Horsten, L., and H. Leitgeb. 2009. How abstraction works. In *Reduction – Abstraction – Analysis*, ed. A. Hieke and H. Leitgeb, vol. 11, 217–226. Berlin: Ontos Verlag.

Lakatos, I. 1976. A renaissance of empiricism in the recent philosophy of mathematics. *British Journal for the Philosophy of Science* 27(3): 201–223.

Lewis, D.K. 1986. *On the Plurality of Worlds*. Oxford: Wiley-Blackwell.

Lewis, D.K. 1991. *Parts of Classes*. Oxford: Blackwell.

Linnebo, O. 2009. Platonism in the philosophy of mathematics. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta. https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=platonism-mathematics.

Linnebo, O. 2018. *Thin Objects*. Oxford: Oxford University Press.

Lowe, E.J. 1995. The metaphysics of abstract objects. *Journal of Philosophy* 92(10): 509–524.

Mancosu, P. 2016. *Abstraction and Infinity*. Oxford: Oxford University Press.

Pettigrew, R. 2008. Platonism and aristotelianism in mathematics. *Philosophia Mathematica* 16(3): 310–332.

Resnik, M.D. 1997. *Mathematics as a Science of Patterns*. Oxford: Oxford University Press.

Rosen, G. 2010. Metaphysical dependence: Grounding and reduction. In *Modality: Metaphysics, Logic, and Epistemology*, ed. B. Hale and A. Hoffmann, 109–136. Oxford: Oxford University Press.

Rosen, G. 2011. The reality of mathematical objects. In *Meaning in Mathematics*, ed. J. Polkinghorne. Oxford: Oxford University Press.

Rosen, G. 2016. Mathematics and metaphysical naturalism. In *The Blackwell Companion to Naturalism*, ed. K. Clark, 277–288. Oxford: Wiley Blackwell.

Schaffer, J. 2009. On what grounds what. In *Metametaphysics: New Essays on the Foundations of Ontology*, ed. D. Manley, D.J. Chalmers, R. Wasserman, 347–383. Oxford: Oxford University Press.

Schaffer, J. 2015. What not to multiply without necessity. *Australasian Journal of Philosophy* 93(4):644–664.

Schwartzkopff, R. 2011. Numbers as ontologically dependent objects hume's principle revisited. *Grazer Philosophische Studien* 82(1): 353–373.

Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.

Wilson, J.M. 2014. No work for a theory of grounding. *Inquiry: An Interdisciplinary Journal of Philosophy* 57(5–6): 535–579.

Chapter 3
Hofweber's Nominalist Naturalism

Eric Snyder, Richard Samuels, and Stewart Shapiro

Abstract In this paper, we outline and critically evaluate Thomas Hofweber's solution to a semantic puzzle he calls Frege's Other Puzzle. After sketching the Puzzle and two traditional responses to it—the Substantival Strategy and the Adjectival Strategy—we outline Hofweber's proposed version of Adjectivalism. We argue that two key components—the syntactic and semantic components—of Hofweber's analysis both suffer from serious empirical difficulties. Ultimately, this suggests that an altogether different solution to Frege's Other Puzzle is required.

Keywords Thomas Hofweber · Number words · Frege's other puzzle · The substantival strategy and the adjectival strategy

3.1 Introduction

This paper is part of a larger project in which we develop an empirically informed, methodologically naturalistic philosophy of mathematics. Our primary concern is with the natural numbers of basic arithmetic, and the idea that empirical results from linguistics, psychology, and cognitive neuroscience may shed light on their nature and our knowledge of them. Our basic conviction is that such a methodologically naturalistic approach can help illuminate traditional core questions that preoccupy

This publication was funded by LMU Munich's Institutional Strategy LMU excellent within the framework of the German Excellence Initiative.

E. Snyder
LMU, Munich, Germany

Ashoka University, India

R. Samuels · S. Shapiro (✉)
Ohio State University, Columbus, OH, USA
e-mail: samuels.58@osu.edu; shapiro.4@osu.edu

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_3

philosophers of mathematics: Do numbers exist, and if so, what are they like? Can we have mathematical knowledge, and if so, how? Indeed, we maintain that such methods can help where more traditional, a priori methodologies cannot.

Against the backdrop of this larger project, the present paper serves two main purposes. First, we provide a detailed exploration of one influential, methodologically naturalistic project of the sort that we seek to pursue: Thomas Hofweber's (2005, 2007, 2016) analysis of number talk and thought, and his defense of nominalism based on that analysis. In doing so, we pay special attention to the manner in which Hofweber's account – in contrast to more traditional, a priori approaches – relies heavily on linguistic considerations in support of this controversial ontological thesis.

Second, we argue that Hofweber's attempt to recruit such considerations to this end is unsuccessful. Hofweber's account is presented as a series of empirical hypotheses, regarding ordinary number-related talk and thought. As such, it is both appropriate – and charitable – to assess his view by the same standards operative in empirical research more generally, including linguistics. Like all local empirical hypotheses, Hofweber's hypotheses should be assessed (among other things) in terms of their ability to generate accurate predictions, and the extent to which they cohere with more basic background theory. By these standards, however, Hofweber's proposal performs poorly. Specifically, insofar as those hypotheses make concrete empirical predictions, they appear to be largely incorrect. In other cases, however, it is not clear whether Hofweber's hypotheses make concrete predictions, or how they may cohere with contemporary linguistic theory. Ultimately, the upshot will be that the best available linguistic evidence does not support nominalism, contra Hofweber.

The rest of the paper proceeds as follows. In Sect. 3.2, we sketch a linguistic puzzle, known as Frege's Other Puzzle, which Hofweber's analysis is primarily designed to solve. We explain how the Puzzle arises, along with two popular philosophical strategies for solving it. In Sect. 3.3, we outline Hofweber's solution, breaking the analysis developed into two components: a syntactic component and a semantic component. In Sect. 3.4, we criticize both components, arguing that neither stand up to empirical scrutiny. We conclude the paper in Sect. 3.5, where we summarize our conclusions and tease out some broader implications for methodologically naturalistic approaches to the philosophy of mathematics.

3.2 Frege's Other Puzzle

Hofweber's analysis is framed largely around a certain linguistic puzzle. In the *Grundlagen*, Frege (1884, §57) notes that number expressions such as 'four' are used in two importantly different ways:

> Since what concerns us here is to define a concept of number that is useful for science, we should not be put off by the attributive form in which number also appears in our everyday use of language. This can always be avoided. For example, the proposition 'Jupiter has four

> moons' can be converted into 'The number of Jupiter's moons is four'. Here the 'is' should not be taken as a mere copula ... Here 'is' has the sense of 'is equal to', 'is the same as' ... We thus have an equation that asserts that the expression 'the number of Jupiter's moons' designates the same object as the word 'four'.

Specifically, 'four' has an "attributive form" witnessed in (1a), and an apparently referential form witnessed in (1b).

(1) a. Jupiter has four moons.
b. The number of Jupiter's moons is four.

On its face, the function of 'four' in (1a) is to count the collection of moons belonging to Jupiter. In this respect, 'four' resembles non-referential expressions like the adjective 'large' or the determiner 'no' in (2):

(2) Jupiter has large/no moons.

On the other hand, (1b) looks like a prototypical identity statement. As such, it apparently involves singular terms, namely 'the number of Jupiter's moons' and 'four'. In this respect, 'four' in (1b) resembles the name 'Wagner' in (3), due to Hofweber (2007).

(3) The composer of *Tannhäuser* is Wagner.

This suggests that 'four' in (1b) is a numeral, or a name of a number.

At the same time, there are clear semantic differences between attributive 'four' and the numeral 'four'. For example, numerals require singular morphology, whereas attributive 'four' requires plural morphology.

(4) a. Which one of these three numbers is even? [Let's see. Three isn't, and five isn't ...] Four {is/??are}.
b. How many of these eight numbers are even? [Let's see. Two is, and six is ...] Four {??is/are}.

Also, while attributive 'four' is typically acceptable with modifiers like 'exactly' and 'almost', the numeral 'four' is not.

(5) a. {Four/??Almost four} is an even number.
b. {Four/Almost four} children clapped.

Furthermore, numerals license entailments like (6a), but attributive 'four' does not.[1]

(6) a. Mary divided four by two yesterday morning $\vDash$ Mary divided four by two yesterday
b. Mary cooked exactly four eggs yesterday morning $\nvDash$ Mary cooked exactly four eggs yesterday

[1] This follows a referentiality test due originally to Kratzer and Heim (1998).

All of this suggests that 'four' serves different, and indeed incompatible, semantic functions: counting collections and naming numbers.

The above is puzzling because, as Felka (2014) notes, it seems that different occurrences of the same expression in semantically equivalent statements ought to serve the same semantic function. Consider the names 'John' and 'Mary' in (7a) and (7b), for instance.

(7) a. John saw Mary at the mall.
b. Mary was seen by John at the mall.
c. John saw a Mary at the mall.

Plausibly, (7a) and (7b) are equivalent because the names serve the same semantic function in those examples, namely to refer. On the other hand, neither (7a) nor (7b) is equivalent to (7c), where 'Mary' is being used instead as a predicate.

This leads to what Hofweber (2005) calls *Frege's Other Puzzle*, which consists of the following four seemingly plausible, but jointly inconsistent, premises.

(FOP1) (1a) and (1b) are semantically equivalent.
(FOP2) The different occurrences of 'four' in (1a) and (1b) are witness to the same expression, namely 'four'.
(FOP3) The different occurrences of 'four' in (1a) and (1b) serve different semantic functions.
(FOP4) Different occurrences of an expression occurring in semantically equivalent statements serve the same semantic function.

FOP1 is taken for granted by everyone in the relevant debate. It is possible, though ultimately unsatisfactory, to deny this premise, however. For example, one could point out that while the follow up to (8a) is perfectly consistent, the follow up to (8b) seems contradictory.

(8) a. Jupiter has four moons. In fact, the number of Jupiter's moons is *sixty-two*.
b. ?? The number of Jupiter's moons is four. In fact, the number of Jupiter's moons is *sixty-two*.

Based on facts like (8a), Horn (1972) argues that attributive uses have *lower-bounded* truth-conditions, so that (1a) is true if Jupiter has at least four moons. Conversely, (8b) might be taken to show that (1b) has *two-sided* truth-conditions, and so is true if instead Jupiter has exactly four moons. If so, then FOP1 would be false. Call this *the Non-Equivalence Strategy*.

The obvious problem with the Non-Equivalence Strategy is that even if we grant that facts like (8) demonstrate the non-equivalence of (1a,b), we can easily reformulate the Puzzle by substituting (9) for (1b) in the original formulation.

(9) Jupiter has exactly four moons.

The follow up to (10) sounds just as contradictory as that of (8b),

(10) ?? Jupiter has exactly four moons. In fact, Jupiter has sixty-two moons.

and yet the Non-Equivalence Strategy would be inapplicable to this new formulation.

It is also possible, though ultimately untenable, to deny FOP2, thus resulting in what we might call *the Homonym Strategy*. According to it, the different occurrences of 'four' in (1a) and (1b) are witnesses to altogether different expressions, ones which just happen to be spelled and pronounced alike. In general, we do not expect homonyms like the noun 'fire' and the verb 'fire' to be acceptably intersubstitutable.

(11) a. The rapid oxidation of combustible materials is fire.
b. Let's {fire/??the rapid oxidation of combustible materials} John.

Similarly, Hofweber notes that substituting 'the number of Jupiter's moons' for 'four' in (1a) leads to unacceptability despite (1b) appearing to establish their coreferentiality.

(12) Jupiter has {four/??the number of Jupiter's moons} moons.

However, because homonyms are typically spelled and pronounced alike as a matter of historical accident, we do not expect their meanings to be related. Thus, the problem with the Homonym Strategy is that the occurrences of 'four' in (1a) and (1b) are clearly semantically related; both tell us something about how many moons belong to Jupiter.

3.2.1 Two Strategies of Analysis

Given the failures of the Non-Equivalence Strategy and the Homonym Strategy, it appears that we must reject either FOP3 or FOP4. It turns out that nearly all approaches within the philosophical literature deny the former, including Frege (1884), Wright (1983), Hodes (1984), Hofweber (2005, 2007, 2016), Moltmann (2013), and Felka (2014). In fact, denying FOP3 is the hallmark of two opposing positions Dummett (1991, p. 99) dubs *the Substantival Strategy* and *the Adjectival Strategy*:

> Number-words occur in two forms: as adjectives, as in ascriptions of number, and as nouns, as in most number-theoretic propositions. When they function as nouns, they are singular terms, not admitting of the plural; Frege tacitly assumes that any sentence in which they occur as adjectives may be transformed either into an ascription of number ... or into a more complex sentence containing an ascription of number as a constituent part. Plainly, any analysis must display the connection between these two uses ... Evidently, there are two strategies. We may first explain the adjectival use of number-words, and then explain the corresponding numerical terms by reference to it: this we may call the adjectival strategy. Or, conversely, we may explain the use of numerals as singular terms, and then explain the corresponding number-adjectives by reference to it; this we may call the substantival strategy.

According to the Substantival Strategy, or *Substantivalism*, both occurrences of 'four' in (1a,b) are in fact numerals, and the apparently non-referential use witnessed in (1a) is to be explained in terms of the genuinely referential use witnessed in (1b). In contrast, according to the Adjectival Strategy, or *Adjectivalism*, both occurrences of 'four' in (1a,b) are either adjectives or determiners,[2] and the apparently referential use witnessed in (1b) is to be explained in terms of the genuinely non-referential use witnessed in (1a).

The most well-known defender of Substantivalism was, of course, Frege (1884). His primary interest was in developing an ideal logical language suitable for science. In such a language, the sole semantic function of a number expression would be to refer to numbers. Thus, non-referential uses of number expressions in natural language are misleading with respect to their ideal semantic function. Consequently, Frege proposes "converting" the attributive form witnessed in (1a) into the singular term witnessed in (1b). To do so, he first proposes paraphrasing (1a) as (1b), and then (equivalently) analyzes the latter as (13).

(13) $\#[\lambda x.\ \text{moon-of-Jupiter}(x)] = 4$

Here, '#' is a cardinality-function mapping a concept Φ to a natural number n representing how many objects fall under Φ. Thus, (13) is an identity statement: it equates a certain number, namely the number of moons belonging to Juptier, with the natural number referenced by the numeral '4', namely four. Thus, on Frege's proposal, (1a) and (1b) should both be analyzed as identity statements, at least for the purposes of an ideal logical language.

Now, as stated, Frege's Other Puzzle is a puzzle about *natural language*: How can one and the same expression serve seemingly different semantic functions in equivalent statements? On the other hand, Frege's analysis was not intended to be a piece of natural language semantics. Rather, as the above quotation from *Grundlagen* §57 makes clear, his primary objective was to "define a concept of number that is useful for science". Thus, the question arises as to whether Substantivalism might be viewed as an independently viable strategy for rejecting FOP3.

Indeed, something like this appears to be rhetorically suggested by Crispin Wright (1983). Speaking of Frege's example *abstraction principle* in (14),

(14) $\forall l_1.\ \forall l_2.\ D(l_1) = D(l_2) \leftrightarrow l_1 \sim l_2$
(For any lines l_1 and l_2, the direction of l_1 is identical to the direction of l_2 just in case l_1 and l_2 are parallel)

Wright (1983, p. 31–32) says the following:

[2] The label "Adjectivalism" is due to Dummett (1991). It is somewhat unfortunate, however, because it suggests that what Frege calls "attributive uses" like (1a) must be adjectives. However, the intended view is that "attributive uses" are *non-referential* expressions, and this is consistent with 'four' in (1a) being an adjective or a determiner. Despite this, we follow the literature in retaining the label "Adjectivalism".

> The reductionist idea was that since the right-hand contains no apparent direction-denoting singular term, we can take it that the apparent reference to a direction on the left-hand side is mere surface grammar, a misleading nuance. But why should we not turn that way of looking at things on its head? What is there to prevent us saying that, since the left-hand side does contain an expression referring to a direction, it is the apparent lack of reference to a direction on the right-hand side which is potentially misleading, or 'mere surface grammar'? ... Why should it not be possible for a sentence containing no isolatable part which refers to a particular object nevertheless achieve, as a whole, a reference to that object, as is attested by the fact that it is equivalent to a sentence in which such a reference is explicit?

Wright's suggestion appears to be that although (15b) but not (15a) contains explicit singular terms referring to directions, because those statements are *equivalent*, we may nevertheless analyze 'line l_1' and 'line l_2' in (15a) as singular terms referring to directions.

(15) a. Line l_1 is parallel to line l_2.
b. The direction of l_1 is identical to the direction of l_2.

Applying similar reasoning to (1a,b), although 'four' in (1a) appears to serve a non-referential semantic function, because (1a,b) are *equivalent*, we may nevertheless analyze 'four' in (1a) as a singular term referring to a number.

However, the obvious problem with this proposal is that because equivalence is *symmetric*, the equivalence of (1a,b) will not alone substantiate Substantivalism. Indeed, a similar point is made by Dummett (1991, p. 109), while criticizing Frege's Substantivalism:

> If it is legitimate for analysis so to violate surface appearance as to find in sentences containing a number-adjective a disguised reference to a number considered as an object, it would necessarily be equally legitimate, if it were possible, to construe number-theoretic sentences as only appearing to contain singular terms for numbers, but as representable, under a correct analysis of their hidden underlying structure, by sentences in which number-words occurred adjectivally... If the appeal to surface form, in sentences of natural language, is not decisive, then it cannot be decisive, either, when applied to sentences of number theory. Frege has merely expressed a preference for the substantival strategy, and indicated a means of carrying it out.

The same criticism applies to Wright's rhetorical suggestion: if we are allowed to ignore surface syntax and analyze the apparent adjective or determiner 'four' in (1a) as a genuine singular term in virtue of the equivalence of (1a,b), then it should be equally legitimate to ignore surface syntax and analyze the apparent numeral in 'four' in (1b) as a non-referential adjective or determiner, thus vindicating Adjectivalism.

In more recent times, Adjectivalism has become by far the more popular solution to Frege's Other Puzzle. Although there are different versions of the strategy available, perhaps the most influential is the one articulated and defended by Hofweber. As we will see, Hofweber's solution has potentially far reaching consequences not merely for the meanings of number expressions, but also for ontology. In the next section, we will outline Hofweber's Adjectivalism, along with its significance for issues central to the philosophy of mathematics.

3.3 Hofweber's Adjectivalism

The version of Adjectivalism defended by Hofweber (2005, 2007, 2016) is complex, consisting of several (controversial) theses. For exegetical clarity, it can be factored it into three major components: a syntactic component, a semantic component, and a cognitive component. Since our primary concern here is with the linguistic aspects of Hofweber's analysis, we will be less concerned overall with the cognitive theses Hofweber puts forward. In what follows, we will sketch these linguistic theses, the solution they recommend to Frege's Other Puzzle, and its implications for the ontology of numbers.

3.3.1 The Syntactic Component: Determiners, Extraction, and Focus Effects

The key semantic fact about natural language determiners is that they cannot function referentially. No empirically respectable semantics would claim that 'no', for instance, can refer to an object. Rather, determiners combine with nouns like 'moon(s)' to form quantificational phrases such as 'no moons', denoting second-order properties (or *generalized quantifiers*).

(16) Jupiter has no/some moons.

So, if 'four' in (1a) is a determiner,

(1) a. Jupiter has four moons.

then it too must function non-referentially. Thus, Hofweber's first key linguistic contention is that 'four' in (1a) is in fact a determiner, one which has a meaning given within *Generalized Quantifier Theory* (GQT; Barwise and Cooper (1981)). On one variation, this is given in (17):[3]

(17) [[four]] = {<*S*,*S*'>: *S*,*S*' ⊆ *U* and |S ∩ S| = 4}
('four' denotes pairs of sets *S* and *S*' such that *S* and *S*' are subsets of the domain *U* and the cardinality of the intersection of *S* and *S*' is exactly four)

[3] (17) is in fact the denotation of 'four' assumed by Breheny (2008). In GQT, cardinal determiners are actually given *lower-bounded* truth conditions, so that 'four' denotes a relation between sets whose intersection has a cardinality of *at least* four:

(i) [[four]] = {<*S*,*S*'>: *S*,*S*' ⊆ *U* and |S ∩ S| ≥ 4}

The reason for adopting a two-sided analysis instead will become apparent in the next section, when we consider paraphrases of basic arithmetic equations like 'three and two is five'.

According to (17), the determiner 'four' denotes a relation between sets whose intersection has a cardinality of exactly four. As such, 'four' in (1a) is thus a prototypical *non*-referential expression. Indeed, as noted just above, expressions of this type not only typically fail to function referentially, they *cannot* function referentially.[4]

Hofweber's primary reason for thinking that 'four' in (1a) is a determiner, as opposed to an adjective, appears to be the undoubted success of GQT. GQT is the predominant analysis of natural language quantification within linguistic semantics, thanks in large part to its ability to state and predict various linguistic universals, specifically generalizations about possible determiner meanings across languages. Thus, Hofweber (2007, p. 3–4) says:

> In contemporary natural-language semantics the uses of 'four' as in [(1a)] are pretty well understood, and 'four' is usually considered to be a determiner, an expression of the same kind as 'some', 'many', and 'all'. Such expressions are not disguised referring terms.

Indeed, if 'four' in (1a) is a determiner, then GQT's success provides an excellent reason for thinking that it expresses a relation between sets.

Hofweber's second key linguistic contention is that 'four' in (1b) *is the very same* quantificational determiner witnessed in (1a).

(1) b. The number of Jupiter's moons is four.

According to Hofweber (2005, p. 211), this is due to what he calls *extraction*.

> In Hofweber [2007], I argue that this focus effect can't be explained if one thinks that [(1b)] is both syntactically and semantically an identity statement with two (semantically) singular terms. But it can be explained if [(1b)] has a different syntactic structure, one that results from extracting the determiner and placing it in an unusual position that has a focus effect as a result. Thus, in [(1b)] 'four' is a determiner that has been "moved" out of its usual position.

The idea appears to be that through "extraction", 'four' in (1a) gets "moved" from its "usual [determiner] position", thereby "placing it in an unusual [post-copular] position" in (1b). Crucially, and despite this, 'four' in (1b) retains its semantic function as a non-referential *determiner*. To quote Hofweber (2005, p. 211): "The word 'four' is the same in [(1a)]) and [(1b)]." Consequently, 'four' in both (1a,b) denotes a property of sets, *not* a number.

Hofweber's main source of evidence for "extraction" concerns so-called *focus effects* witnessed in examples like (18).

(18) a. Johan likes soccer.
b. What Johan likes is soccer.
c. It is Johan who likes soccer.

Whereas (18a) is acceptable in response to both 'Who likes soccer?' and 'Which sport does Johan like?', (18b) is only acceptable in response to the latter, while

[4] See Landman (2003).

(18c) is only acceptable in response to the former. Contrast this with prototypical identity statements like (19), which apparently do not give rise to focus effects.

(19) Cicero is Tully.

Indeed, (19) is perfectly fine in response to both 'Who is Tully?' and 'Who is Cicero?'.

In contrast, (1b) does apparently display focus effects: while (1a) is acceptable in response to both 'Which planet has four moons?' and 'What belongs to Jupiter?', (1b) is only acceptable in response to the former. What this shows, according to Hofweber, is that (1b) is not a genuine identity statement, contra Frege (1884). Moreover, it is indirect evidence that (1b) results from "extraction" since, if it were an identity statement, we would expect to see no focus effects.

3.3.2 The Semantic Component: Numerals and Semantically Bare Determiners

To summarize, according to Hofweber, 'four' in (1b) is the same non-referential determiner witnessed in (1a), thanks to "extraction". As such, the truth of neither (1a) nor (1b) implies the existence of a number, no more so than (16) does.

(16) Jupiter has no/some moons.

However, "extraction" is a construction-specific syntactic operation, presumably: it applies to sentences broadly having the structure of (1a), and returns sentences broadly having the structure of (1b). As such, it is not operative in numerical equations like (20).

(20) Three and two is five.

After all, it is hard to see how (20) could result from anything similar to (1a), where the numerals 'three', 'two', and 'five' feature originally as determiners. But then there is no obvious reason for thinking that the numerals in (20) are non-referential expressions. In other words, it would appear that (20) straightforwardly entails the existence of numbers.

To this end, Hofweber distinguishes between two kinds of *bare determiners*, or determiners occurring without overt accompanying nouns, such as 'most' in (21).

(21) How many boys kicked the ball? Most kicked the ball.

Although 'most' does not occur explicitly restricted by the noun 'boys' in (21), it is implicitly understood that way. In other words, the continuation of (21) is interpreted as 'Most boys . . . '. Contrast this with 'most' in (22), where there is no antecedent noun available.

(22) Most is/are more than none.

Rather than claiming something about most boys, or most people, or whatever, (22) is intended to be *generic*: whatever it is that we're talking about, most is more than none. Thus, Hofweber calls determiners like 'most' in (22) *semantically bare determiners*.

Hofweber's third key linguistic contention is that the number expressions in (22) are really semantically bare determiners, not genuine names of numbers. Put differently, (22) has something like the logical form informally suggested in (23), where *X* is a noun phrase restricting the determiners 'three', 'two', and 'five', and 'GEN' is a genericity operation.

(23) GEN: [three *X* and two (more) *X* are five *X*]
(In general, three things and two (more) things are five things)

Hofweber's primary piece of linguistic evidence for (23) is that arithmetic equations can be parsed in two ways, namely in the singular or in the plural, similar to (22).

(24) Three and two is/are five.

Furthermore, (24) also resembles (22) in that both are entirely general: no matter what we are talking about, three and two are five, just as most are more than none. As a result, despite surface syntactic appearances, (24) does not involve relating two first-order objects (3 and 2) to a third first-order object (5), through a first-order operation (+). Rather, it actually involves *counting* objects, though in an entirely general way.

This raises another question, however: What guarantees that things (the X's) being counted in (24) do not overlap. This is crucial to getting the truth-conditions for (24) correct, of course: if A = {a,b,c}, B = {a,b}, and C = {d,e}, then $|A \cap B| = 2$, $|A \cup B| = 3$, and $|A \cup C| = 5$. So, what guarantees that (22) behaves like $|A \cup C|$, rather than $|A \cap B|$ or $|A \cup B|$? To this end, Hofweber appeals to a well known distinction between *cumulative* (or 'non-boolean') conjunction and *propositional* (or 'boolean') conjunction. Examples of the latter include (25a-c), while examples of the former include (26a-c), due to Krifka (1999).

(25) a. John and Mary slept.
b. Mary sang and danced.
c. This cocktail is cheap and refreshing.

(26) a. John and Mary met at the mall.
b. This concoction is beer and lemonade.
c. That flag is entirely green and white.

(25a–c) can all be paraphrased as the conjunction of two propositions. For example, (25a) can be paraphrased as 'John slept and Mary slept'. In contrast (26a) cannot mean that John met at the mall, and also Mary met at the mall, just as (26b) cannot

mean that this concoction is beer, and also this concoction is lemonade. What (26a–c) show is that cumulative conjunction can coordinate expressions of different semantic types – names, predicates, and modifiers.

Thus, Hofweber's fourth linguistic contention is that 'and' in (22) and (24) is *cumulative* conjunction involving semantically bare determiners, where non-overlap is guaranteed through "ellipsis, or a pragmatic mechanism, or a form of "free enrichment," or something else" (Hofweber (2005, p. 193)). Thus, Hofweber likens (24) to (27).

(27) She only had an apple and dessert.

Normally, an utterance of (27) would be judged misleading if she happened to have only an apple, even though apples may serve as perfectly fine desserts. Presumably, according to Hofweber, this too is a function of "ellipsis, or a pragmatic mechanism, or a form of "free enrichment," or something else". The important point is that just as an utterance of (27) apparently presupposes non-overlapping extensions for 'apple' and 'dessert', an utterance of (24) apparently presupposes non-overlapping extensions of 'three *X*s' and 'two *X*s'.

3.3.3 Frege's Other Puzzle and the Consequences for Ontology

Given that Hofweber defends a version of Adjectivalism, it is hardly surprising that the premise he denies in Frege's Other Puzzle is FOP3.

> The different occurrences of 'four' in (1a) and (1b) serve different semantic functions.

Specifically, despite 'four' occurring as a determiner in (1a) and as a name in (1b), it serves the same non-referential semantic function of a determiner in both.

(1) a. Jupiter has four moons.
b. The number of Jupiter's moons is four.

As a result, (1b) does not entail the existence of a number.

Contrast this with Frege's analysis,[5] where (1b) immediately implies the existence of a number thanks to the referential function of the numeral 'four'. That is, (28a) entails (28b), paraphrased in English as (28c).

(28) a. $\#[\lambda x.\ \text{moon-of-Jupiter}(x)] = 4$
b. $\exists n.\ \#[\lambda x.\ \text{moon-of-Jupiter}(x)] = n \wedge n = 4$
c. There is a number which is the number of Jupiter's moons, namely four.

[5] The same applies to Wright (1983) and Hale (1987).

Since (28c) seemingly wears its ontological commitment to numbers on its sleeves, and since (1a) entails (1b), Frege's analysis apparently implies that in virtue of successfully counting some moons, thereby establishing the truth of (28a), we can validly infer that numbers exist, i.e. (28b). This is puzzling, as the question of whether numbers exist is a longstanding, difficult question central to the philosophy of mathematics. Thus, it would be surprising if an answer to that question could be so easily obtained. Accordingly, this is known in the literature as *the Easy Argument for Numbers*.[6]

Hofweber (2007)'s solution appeals to the same Adjectivalist analysis responsible for debunking Frege's Other Puzzle: because 'four' in (1b) is a non-referential determiner, we cannot infer from it that a number exists, at least not in a substantial sense relevant to ontology. Ultimately, and more generally, Hofweber's view is that *no* apparently referential use of number expression is in fact referential, including their use in arithmetic statements, such as (24).

(24) Three and two is/are five.

Thus, in the end, Hofweber defends a version of what Dummett (1991) calls *the Radical Adjectival Strategy*: *no* occurrence of e.g. 'four' or '4' is a genuine singular term.[7] Consequently, not only does our ordinary talk of counting moons fail to entail an ontology of numbers, but so also does the mathematician's talk of the number four being even.

In summary, Hofweber's Adjectivalism may thus be viewed as a sustained defense of nominalism with respect to the natural numbers. All apparent reference to numbers is just that – apparent. Upon further linguistic investigation, we discover that explaining arithmetic truths does not require positing numbers. That's because, despite surface appearances, arithmetic discourse is not about numbers as abstract objects and the various properties those objects may have, but rather an elaborate form of counting, something we all learned to do as children. What's more, arithmetic discourse is *true* – indeed objectively and necessarily so – in virtue of the meanings of the bare numerical determiners involved. Thus, unlike with various versions of *fictionalism*,[8] Hofweber's Adjectivalism is not an error theory with respect to number talk.

Of course, the strength of Hofweber's defense of nominalism depends wholly on the empirical adequacy of the analysis proposed. In the next section, we will sketch objections to the two components of the analysis considered here – the syntactic component and the semantic component. Ultimately, we will argue that neither survives empirical scrutiny.

[6] See e.g. Balcerak-Jackson (2013) and Snyder (2017).

[7] The details are complex and beyond the scope of a single paper. But see Hofweber (2016).

[8] Roughly, fictionalism is the view that numerals in arithmetic discourse genuinely have the function of naming numbers, but since numbers do not exist, all arithmetic discourse is either false or else involves widespread presupposition failure. See e.g. Hodes (1984), Yablo (2005), and Leng (2010).

3.4 Problems with Hofweber's Adjectivalism

Despite its influence, Hofweber's Adjectivalism has received a fair amount of criticism in the philosophical literature. This has focused largely on the syntactic component of Hofweber's analysis, specifically "extraction" and the evidence purporting to motivate it. In this section, we will consider those objections, while also developing novel objections to the comparatively neglected semantic component of Hofweber's analysis. The upshot will be that Hofweber's key linguistic theses highlighted in Sect. 3.3 are empirically problematic.

3.4.1 Problems with Extraction

Much of the extant criticism of Hofweber's Adjectivalism revolves around "extraction", i.e. the linguistic mechanism responsible for "moving" 'four' from its position in (1a) to its position in (1b). Recall the quote from Hofweber (2005, p. 211):

> [(1b)] has a different syntactic structure, one that results from extracting the determiner and placing it in an unusual position that has a focus effect as a result. Thus, in [(1b)] 'four' is a determiner that has been "moved" out of its usual position.

It is natural to interpret this talk of "movement" as an instance of the same kind of "movement" familiar from transformational theories of syntax (e.g. transformational grammar, government and binding theory, and minimalist syntax). (29) provides a prototypical illustratration, known as "extraposition", where underlining indicates the expression "moved", and the blank indicates the position out of which "movement" is assumed to occur.

(29) a. Something that we weren't expecting happened.
b. Something ___ happened that we weren't expecting.

As a result, 'that we weren't expecting' in (29b) becomes *focused*, much like post-copular 'four' in (1b) on Hofweber's analysis. It should thus be unsurprising that some of Hofweber's detractors, chiefly Brendan Balcerak-Jackson (2013) and Friederike Moltmann (2013), have interpreted "extraction" as a transformational mechanism responsible for "rearranging" the syntactic material in (1a), ultimately resulting in (1b).

The problem, according to these detractors, is that the actual syntactic principles or operations that would be required to do this kind of "rearranging" would not be recognized by contemporary transformational theories, and their postulation would be highly dubious. For one thing, unlike with (29), (1b) clearly contains material missing in (1a): 'the', 'number', 'of', and '-'s'. Conversely, there is material contained in (1a) that is missing in (1b): 'has'. Even if there were "movement" of the parts of (1a), no other known transformational mechanism would also delete and add material in the manner required. Rather, as Balcerak-Jackson notes, it would seem far more plausible to hold that (1a,b) are simply *different sentences*, with

(1b) attempting to paraphrase (1a). Yet without (1b) resulting from some kind of "rearrangement" of (1a), there would be no guarantee that post-copular 'four' in (1b) is *the same* non-referential determiner (1a), thereby undermining Hofweber's case for some version of Adjectivalism.

In response, Hofweber (2014) accuses these detractors of misinterpreting "extraction", by assuming that it must involve some kind of transformational "movement". To quote Hofweber (2014, p. 264):

> But I made no such proposal. I never talk about transformation rules, or deriving [(1b)] from [(1a)] via some mysterious sentence level transformation. In fact, 'transform' or 'transformation' don't even appear in my article.

To help clarify, Hofweber further distinguishes between two possible interpretations of "extraction", one involving what he calls "displacement", and the other involving what he calls "transformation". To continue the quote:

> To bring out the difference, we can say that 'extraction' could be understood either as displacement or as transformation. Displacement occurs when a phrase appears in a position contrary to where it naturally belongs, that is, contrary to its canonical position. This is still metaphorical, of course, but at least talk of displacement rather than extraction might suggest less that this is to be understood as sentence level transformation. Transformation occurs when one sentence gets turned into another, via some syntactic rules. I proposed that in [(1b)], but not in [(1a)], 'four' is displaced and as a result we can see, in outline, why [(1b)] has a focus effect, while [(1a)] does not. Balcerak Jackson instead takes me to propose that [(1a)] gets transformed into [(1b)]. All that is needed for the argument, however, is displacement, not transformation.

Thus, according to Hofweber, the Balcerak-Jackson/Moltmann criticism is ultimately a straw man, requiring a "transformation"-based interpretation of "extraction", rather than a "displacement"-based interpretation.

Suppose so. The obvious question now becomes: What exactly is "displacement", and what does it have to do with 'four' in (1a) getting "moved" into post-copular position in (1b)? Unfortunately, Hofweber has little to offer in response to these questions. To quote Hofweber (2014, p. 265) again:

> In Hofweber (2007) I did not propose any particular view of how the syntax for the relevant examples was supposed to work more precisely. I made no proposal about the precise syntactic structure of [(1b)], nor about the relationship between focus and syntax in general, nor did I endorse a particular framework in syntactic theory. I don't say this proudly, I wish I had such views to offer. But the argument that [(1b)] is not an identity statement is rather neutral with respect to the more precise syntactic mechanisms that underlie all this. It is motivated more by the data for a theory than the theory itself. It relied on a notion of extraction/displacement that was metaphorical, but clear for many cases, its connection to syntactic focus, and the relationship between focus and identity statements, but not any particular syntactic theory, certainly not transformational grammar.

A similar sentiment is expressed in Hofweber (2016, p. 41):

> Talk of "extraction" or "displacement" or "movement" is a theory-neutral metaphor that we don't need to spell out now. What is crucial for us instead is that constructions of this kind give rise to a syntactic focus effect, not how precisely the syntactic connection to focus is to be understood.

So, "displacement", and thus "extraction", is only intended to be a metaphor, not a fleshed out syntactic mechanism situated within the background of a particular syntactic theory, including transformational theories.

In some ways, it is understandable that Hofweber might want to back off from making any specific proposals regarding how exactly "extraction" is to be understood. After all, doing so might lead to potentially falsifiable empirical predictions, and it might also hold his analysis hostage to cohering with other elements of a background syntactic theory. On the other hand, because Hofweber's solution to Frege's Other Puzzle rests wholly on the *empirical viability* of "extraction", what's required, minimally, is some assurance that this syntactic mechanism, whatever it is, is empirically motivated.

There are two points we'd like to emphasize here in this connection. First, it should be stressed that "movement" of any sort is highly controversial within contemporary syntactic theory. That's because there are *numerous* mainstream syntactic theories whose formulation is grounded principally upon the explicit rejection of "movement", notably "representational theories", including head-driven phrase structure grammars, lexical functional grammars, construction grammars, and most dependency grammars. The latter attempt to explain the same phenomena covered by the postulation of "movement" within transformational theories, but via other means, e.g. feature passing.

Secondly, and largely for this reason, talk of "movement" is typically understood as presupposing some version of transformational syntax. This includes "displacement". To illustrate, consider the following example from Abels (2017), where again underlining indicates the expression "moved", and gaps indicate the "canonical position" from which it is "displaced":

(30) a. (I know that) John will drink absinthe.
b. I know what John will drink ___.
c. Absinthe, John will drink ___.
d. the beverage which John will drink ___

As Abels explains, the "canonical" word order for English sentences, as illustrated in (30a), is subject-auxiliary-verb-object. In (30b-d), the underlined expression is the object, thus revealing that it does not occur within its "canonical" position – it has been "displaced". This is also presumably the notion of "displacement" Hofweber (2014) has in mind: "a phrase appears in a position contrary to where it naturally belongs, that is, contrary to its canonical position."

If so, then it is very difficult to see how Hofweber's distinction between "displacement" and "transformation" addresses the crux of the Balcerak-Jackson/Moltmann criticism. Hofweber's central thesis is that 'four' in (1b) results from "displacement", and as such is the very same determiner witnessed in (1a). For this to make sense, 'four' needs to be "moved" out of its "canonical position" in (1a) , presumably as the head of a determiner phrase, to post-copular position in (1b). Minimally, then, as with typical cases of "displacement" like (30b-d), we should expect (1a) to *share* much of its syntactic material with (1b), contrary to fact. Thus, independent of any particular version of transformational syntax and

corresponding syntactic principles or operations which might underlie this kind of "movement", it would appear that "displacement" cannot do what Hofweber's Adjectivalism requires of it.

In any case, there would appear to be more direct evidence against "extraction", independent of how it might be spelled out. Generally speaking, we expect syntactic operations to apply to expressions of the same categories. For example, we should presumably be able to replace 'absinthe' in (30a) with any other mass noun, thus leading to a grammatical sentence of the form in (30c). So, if "extraction" is a syntactic operation, then it should apply to all determiners, not just numerical determiners. For example, it should apply to 'no' and 'some' in (31).

(31) Jupiter has {no/some/four} moons.

Yet, as Balcerak-Jackson also points out, the result of applying "extraction" to 'no' and 'some' would be clearly unacceptable:

(32) The number of Jupiter's moons is {??no/??some/four}.

More generally, it appears that *no* uncontroversial determiner can occur in post-copular position of constructions like (1b). Why is this?

In response, Hofweber (2014, p. 266) says the following:

> Balcerak Jackson contends that my account does not explain why similar constructions do not seem to work with other determiners. That is true, my account does not explain this, and neither does, I may add, Balcerak Jackson's own account outlined at the end of his paper. But it is an overstatement that my account makes this "mysterious" (p. 451). The account simply leaves this open, but it is certainly compatible with an explanation that comes form a difference in the syntactic behavior among determiners or adjectives in general. That not all determiners behave the same syntactically is a well-known fact.

Thus, Hofweber's response to this objection looks similar to his response to the previous objection: because "extraction" is not intended to be situated within any particular syntactic theory, it is not intended to offer an explanation of contrasts like (32). Rather, (32) is apparently witness to a more general phenomenon, of which Hofweber unfortunately offers no specific examples, that "not all determiners behave the same syntactically".

The problem with this response, as with the first, is that it does not actually address the argument at issue. The concern is not with whether Hofweber's analysis can explain contrasts like (32), but rather with what it apparently predicts. Specifically, the claim is that Hofweber's analysis seemingly makes a false empirical prediction: all determiners should be subject to "displacement", and yet *no* uncontroversial determiners can acceptably occur in post-copular position, similar to 'four' in (1b). Now, Hofweber's response could be that his analysis does not make this prediction, because it makes *no predictions*, in virtue of not being situated within any particular syntactic theory. But this would ignore the crucial fact about syntactic operations more generally: they apply to expressions of the same category, *independent* of whichever syntactic theory they happen to be embedded within. Thus, independent of which specific syntactic operation is assumed to be responsible

for "displacement" (e.g. "Inner Merge"), Hofweber's analysis appears to make an important, demonstrably false prediction.

Furthermore, note that Balcerak-Jackson's original observation readily extends to numerous further constructions. For example, uncontroversial determiners cannot appear bare in predicative positions more generally:

(33) Jupiter's moons are {??no/??some/four} (in number).

Nor can they occur as the complement of the verb 'number':

(34) Jupiter's moons number {??no/??some/four}.

Nor can they generally be "stacked", i.e. co-occur bare.

(35) All {??no/??some/four} moons of Jupiter are large.

Finally, and perhaps most significantly, determiners cannot occur as *names*.

In contrast, color expressions such as 'green' can occupy these various positions, and they can also apparently function as names.

(36) a. Jupiter has green moons.
b. The color of Jupiter's moons is green.
c. Jupiter's moons are green (in hue).
d. Jupiter's moons are colored green.
e. All green moons of Jupiter are large.
f. Green is a color.
g. The color green is Mary's favorite.

What's more, such expressions are standardly assumed within linguistic semantics to be *adjectives*, and, in any case, they are certainly *not* determiners. Furthermore, their use in constructions like (36f) is also standardly assumed to be referential, so that 'green' in (36f) is a genuine singular term.[9]

Here's a simple argument, then, building on Balcerak-Jackson's original observation. On the one hand, number expressions differ from other uncontroversial determiners in numerous important respects. On the other hand, they pattern exactly like certain adjectives in those same respects. Furthermore, merely announcing that "not all determiners behave the same syntactically" will not suffice to explain these similarities and differences, given their breadth. Rather, it is tempting to conclude based on these observations that Hofweber's analysis rests principally on a syntactic misclassification – number expressions, at least in their "attributive" use, are *adjectives*, not determiners.

It is thus worth noting that according to Hofweber (2016, p. 124), the issue of how exactly number expressions should be syntactically classified is ultimately irrelevant to the success of his solution to Frege's Other Puzzle.

[9] See e.g. Kennedy and McNally (2010), McNally (2011), and McNally and de Swart (2011).

> There is some controversy about whether number words in the relevant uses are determiners, modifiers, or adjectives. This is also an issue which is insignificant for us here . . . What ultimately matters for our discussion is that number words in their determiner use can form complexes . . . and that they are not themselves referring expressions in this use. Whether they are in the end adjectives, determiners, or form a separate class of their own, is secondary.

Presumably, the thought is that because both determiners and adjectives (used attributively or predicatively) function *non*-referentially, so long as 'four' (1b) also functions that way, Hofweber's proposed solution to Frege's Other Puzzle will go through.

However, we have just seen at least some (apparent) adjectives have genuinely referential uses – again, witness (36f). In fact, there are analyses on which 'four' in (1b) arguably has this same referential semantic function, in virtue of the same general semantic operations responsible for rendering 'green' in (36f) a singular term.[10] What's more, (36) is standardly taken to show that one and the same expression ('green') can perform different semantic functions – it can function e.g. as a predicate, a modifier, and as a singular term. Moreover, all "extraction" apparently guarantees is that, to quote Hofweber (2005, p. 211) again, "the word 'four' is the same in both [(1b)] and [(1a)]". If so, then this alone will not guarantee that 'four' in (1b) has the *same semantic function* as four in (1a). Rather, if 'four' in (1a,b) is an adjective, then nothing obviously precludes the possibility that 'four' in (1b) is a genuine singular term, despite being "displaced". In contrast, because it is a *distinguishing feature* of determiners that they cannot function referentially, no such possibility arises if 'four' in (1a) is instead a determiner. It thus appears that the classification of number expressions in their "attributive" use is far more empirically significant than Hofweber recognizes.

3.4.2 Problems with Focus Effects

In addition to "extraction", Hofweber's analysis relies crucially on a number of dubious semantic assumptions. One concerns the role of so-called focus effects. Hofweber's argument, recall, is that because genuine identity statements do not exhibit focus effects, but (1b) does, (1b) cannot be a genuine identity statement. However, Brogaard (2007) points out that (3), which Hofweber (2007) claims to be a genuine identity statement, exhibits similar focus effects.

(3) The composer of *Tannhäuser* is Wagner.

In particular, (3) would be an appropriate answer to the question 'Who composed *Tannhäuser*?' but not 'Who is Wagner?' or 'What did Wagner do?'. Thus, it appears that exhibiting focus effects is insufficient to show that (1b) is not an identity

[10] See Snyder (2017).

statement. In that case, it could well be that (1b) entails the existence of a number, just as Frege (1884)'s analysis suggests.

Indeed, it is worth emphasizing in this connection that even if constructions like (1b) and (3) are not identity sentences, this alone will not establish that post-copular 'four' in (1b) functions non-referentially. In fact, it has been argued by many that constructions like (1b) and (3) are better understood as *specificational sentences*, in the sense of Higgins (1973).[11] Higgins originally distinguished between at least three forms of the English copula, including:

(37)	a. Cicero is Tully.	(equative)
	b. Cicero is bald.	(predicational)
	c. The most famous Roman orator is Cicero.	(specificational)

Equative sentences are prototypical identity statements like (37a), equating the referents of two singular terms. Predicational sentences such as (37b) predicate a property such as being bald of the subject. Finally, specificational sentences such as (37c) specify an individual under a certain description, e.g. the most famous Roman orator.

The semantic motivations for this taxonomy are well known.[12] Moreover, it might be reasonably thought that Hofweber could appeal to that taxonomy not only to explain the apparent focus effects in (1b) and (3), but also to establish a different version of Adjectivalism which does not rely on the dubious syntactic operation of "extraction". In fact, this is broadly the strategy pursued by other Adjectivalists, including Moltmann (2013) and Felka (2014). On these analyses, specificational sentences more generally express question-answer pairs, via ellipsis.[13] For example, the pre-copular material in (37c) expresses an indirect question corresponding to 'Who is the most famous Roman orator?', while the post-copular material expresses an answer to that question, namely 'Cicero is the most famous Roman orator', all through ellipsis.

Similarly, it has been argued that the pre-copular material in (1b) expresses an indirect question corresponding to 'What is the number of Jupiter's moons?', an answer to which is expressed by the post-copular material, namely 'Jupiter has four moons', again through ellipsis. Since the pre-copular material expresses a question, it is little wonder that we see focus effects. After all, by hypothesis the question expressed concerns the cardinality of Jupiter's moons, not what belongs to Jupiter more generally. Better yet, because post-copular 'four' has the same non-referential function witnessed in (1a), (1b) would not entail an ontology of numbers.

However, for this suggestion to succeed, it needs to be that specificational sentences really do express question-answer pairs in virtue of ellipsis. Yet this is not the *only* analysis of specificational sentences available. In fact, there is an alternative analysis on which the copula is systematically ambiguous between equa-

[11] See Moltmann (2013), Felka (2014), and Snyder (2017).

[12] See Mikkelsen (2005).

[13] See Schlenker (2003).

tive, predicational, and specificational meanings.[14] In particular, the specificational copula receives the meaning in (38), where '$\underline{y}$' ranges over *individual concepts*, i.e. functions from worlds to individuals.

(38) $\lambda x.\lambda \underline{y}_{<s,e>}.\lambda w.\ \underline{y}(w) = x$

On this analysis, the pre-copular material in (37c) expresses an individual concept, namely a function from worlds w to whoever is the most famous Roman orator in w, while the post-copular name 'Cicero' is a genuine singular term. Thus, (37c) will be actually true if Cicero is in fact the most famous Roman orator.

Applying the same analysis to (1b) would suggest that the pre-copular material expresses an individual concept, i.e. a function from worlds w to the maximal number of Jupiter's moons in w, while post-copular 'four' functions as a genuine *singular term*. Hence, even if (1b) is not a genuine identity statement, i.e. a copular sentence involving the equative copula, it needn't follow that post-copular 'four' functions non-referentially. In other words, it needn't follow that some version of Adjectivalism is correct. Ultimately, then, focus effects lend no direct support for Adjectivalism of any kind.

3.4.3 Problems with Numerals

Consider (39), where 'four' apparently functions as a numeral, i.e. a name:

(39) Four is an even number.

Clearly, 'is an even number' is a predicate. Given standard semantic assumptions, it should thus be something which either takes 'four' as an argument and returns a truth-value, or else is taken by 'four' as an argument and returns a truth-value. In the first case, 'four' would be a referential-type expression, presumably referring to a number. In the second case, it would function as a generalized quantifier, denoting a set of sets, one of which would include the even numbers. In either case, it would appear that making semantic sense of the truth of (39) requires acknowledging the existence of numbers.

Of course, this realist conclusion might be avoided if 'four' functions instead as a semantically bare determiner, in which case 'four' and 'is an even number' would have different semantic types than their surface syntax suggests. The problem, however, is that determiners without accompanying nouns generally have the wrong semantic type to occupy argument positions, as witnessed by the unacceptability of 'every' in (40a,b).

(40) a. {??Every/Everyone/Every person/Mary} is happy.
b. John loves {??every/everyone/every person/Mary}.

[14] See Partee (1986b) and Romero (2005).

Rather, in order to occupy argument positions, determiners need to combine with a noun like ‘person’ to form a generalized quantifier like ‘every person’.

Hofweber is seemingly aware of this issue. Indeed, speaking about an example similar to (39), Hofweber (2005, p. 209–210) says:

> I will cover only the case of the relationship between sentences like [(1a,b)]. It will not be a general account of the singular-term use of numerals. It will still leave open what is going on in certain other uses of number words as singular terms in statements that are neither singular basic arithmetical equations nor of the same kind as [(1b)].

This is rather surprising, given that Hofweber’s analysis is designed to handle cases like (24), which also apparently involve numerals.

(24) Three and two is/are five.

The latter, recall, is analyzed as (23), so that the apparent numerals in (24) are in fact generically quantified semantically bare determiners.

(23) GEN: [three *X* and two (more) *X* are five *X*]
[In general, three things and two (more) things are five things]

Hofweber’s contention is that because determiners more generally are non-referential expressions, (24) does not entail commitment to numbers. Thus, one might reasonably think that something similar could be said for cases like (39), so that ‘four’ similarly functions as a semantically bare determiner.

However, as Rothstein (2017) observes, this suggestion would make numerous incorrect predictions. First, ‘count’ is ambiguous between two senses, roughly corresponding to what Benacerraf (1965) calls *intransitive counting* and *transitive counting*.[15] These are witnessed respectively in (40a,b), due to Rothstein.

(40) a. I counted to thirteen (??things/??people/??books).
b. I counted thirteen (things/people/books).

Thus, as the labels suggest, transitive ‘count’ requires a direct object, where intransitive ‘count’ does not. Semantically, this suggests that while transitive ‘count’ has an essentially *relational* meaning, intransitive ‘count’ does not. Thus, consider Rothstein’s (41a,b):

(41) a. I counted thirteen. – Thirteen what?
b.?? I counted to thirteen. – Thirteen what?

Secondly, (42a) and (42b) are clearly not synonymous, as (42a) is true but not (42b).

[15] To a first approximation, intransitive counting consists in reciting the numerals in their canonical order –“1, 2, 3,...” In contrast, transitive counting consists in the counting *of* things. That is, when transitively counting we use the numerals to answer ‘how many’-questions, roughly by establishing a one-to-one correspondence between an initial segment of those numerals and a collection of objects being counted.

(42) a. Two is an even prime.
b. Two things are even primes.

Third, numerals and bare determiners differ in their agreement features. Specifically, whereas numerals require singular morphology, bare determiners require plural morphology.

(43) a. Which one of these three numbers is Mary's favorite? Four {is/??are}.
b. How many people are coming to the party? Four {??is/are}.

And the same holds for numerals in comparative constructions, as Rothstein points out.

(44) Four {is/??are} bigger than three.

Finally, Rothstein observes examples like (45), where one number expression clearly modifies another.

(45) Two twos are four, three twos are six.

None of this is to be expected, however, if all numerals are really semantically bare determiners. In that case, for instance, (42a) would entail (42b), contrary to fact. Rothstein (2017, p. 28) thus reasonably concludes: "Together, these data show that... there are cases where a bare cardinal numerical must be a singular term."

If so, and if their most plausible candidate referents are *numbers*, as argued by Hale (1987), then it would appear that the truth of e.g. (39) straightforwardly entails the existence of a number. In other words, despite Hofweber's proposed solution to the Easy Argument involving (1b), which crucially relies on 'four' functioning as a semantically bare determiner, there would appear to be an equally "easy argument" involving (39), for which that solution does not apply.

To be fair, Hofweber does offer an explanation as to why (apparent) numerals like 'four' in (38) at least *appear* to function as genuine singular terms. The explanation appeals to what he dubs *cognitive type-coercion*. In essence, cognitive type-coercion is the cognitive analog of type-shifting. However, whereas type-shifting is typically taken to "coerce" the meanings of natural language expressions, "shifting" their lexical meanings (at least) in the presence of type-mismatches, cognitive type-coercion instead operates exclusively on *mental representations*, within the language of thought. To quote Hofweber (2016, p. 137):

> The process of cognitive type coercion forces a representation to take on a certain form so that a certain cognitive process can operate with this representation. Systematically lowering the type of all expressions (or the mental analogue thereof) is a way of doing this, and the difference between our ability to reason with representations involving low types rather than high types explains why this type lowering occurs in the case of arithmetic.

The basic idea appears to be that because number expressions occurring within arithmetic statements have the complex semantic type of a determiner, they are difficult to semantically process. Consequently, we are forced to "coerce" the

corresponding mental representations in such a way that our reasoning mechanisms can "get a grip".

It is in virtue of this kind of cognitive coercion, apparently, that numerals in arithmetic statements seemingly function as singular terms. To quote Hofweber (2016, p. 137) again:

> Note that according to the cognitive type coercion account we merely change the form of the representation. We do not replace one representation with another one that has a different content. We take the same representation and change its syntactic form so that our reasoning mechanism can operate on it. The content of what is represented remains untouched by this. To put it in terms of the language of thought, we change the syntax of a representation so that our reasoning mechanism can get a grip on these representations. Other than that we leave it the same. And what holds good for mental representations will hold good, *mutatis mutandis*, for their linguistic expression in language. The singular arithmetical statements are the linguistic expression of thoughts involving type lowered mental representations.

It is crucial to recognize that the kind of "coercion" being alluded to here is *not* type-shifting of the more familiar semantic variety.[16] If it were, then numerals occurring within arithmetic statements would need to function as genuine singular terms, thus resulting in a different version of the Easy Argument. Presumably, this is why, according to Hofweber (2016, p. 141), "semantic type coercion [i.e. type-shifting] is the second best attempt to solve Frege's Other Puzzle." Regardless, the claim appears to be that *within the language of thought*, numerals occurring in arithmetic statements like (24), or their cognitive analogs, do function referentially, and this presumably explains why they appear to function referentially in English as well. If so, then perhaps this explanation can be extended to numerals as they figure in arithmetic statements like (39) as well.

We have criticized the notion of cognitive type-coercion and its role within Hofweber's larger nominalist program at length elsewhere.[17] Here, we will limit our discussion to how this might help explain contrasts like (40)–(45). The latter are presented (by a linguist) as *semantic* contrasts, intended to reveal a difference in the semantic function of numerals (referential) and numerical modifiers (non-referential), in English. However, by hypothesis, cognitive type-coercion operates on mental representations within the language of thought, *not* the meanings of English expressions. The claim appears to be that because we cognitively lower the "types" of corresponding representations at least when dealing with arithmetic statements, this explains why (apparent) numerals in English seem to function referentially.

The question here is: How, exactly? As far as we can tell, Hofweber offers no concrete answer. However, perhaps the most obvious answer is that the judgments reported in (40)–(45) do not reflect anything about the *meanings* of the English expressions at all, but rather their cognitive analogs within Mentalese. If so, then a primary question for Hofweber's account, as we see it, is this: What prevents

[16] See especially Partee (1986a).

[17] See Snyder et al. (2021).

all (purported) semantic judgments from likewise reflecting something about Mentalese? After all, presenting contrasts like (40)–(45) is a primary empirical tool available to semanticists, with the presumption being that such contrasts report native speakers' intuitions about the meanings of natural language expressions. So, if such contrasts actually fail to reveal what linguists standardly take them to reveal, then why should this response, if correct, not rob linguistic semantics of its very empirical foundations?

3.4.4 Problems with Semantically Bare Determiners

We have seen that there are problems with analyzing numerals as semantically bare determiners. To this end, it is worth reexamining Hofweber's original justification for claiming that 'four' in (1a) is a determiner in the first place. It relies crucially, recall, on the observation that GQT is the predominant analysis of natural language determiners. Thus, if 'four' in (1a) is also a determiner, then all other things being equal, we ought to assume that it has the non-referential meaning GQT attributes to it. Specifically, we should assume that it has something like the meaning in (17).

(17) [[four]] = {<*S*,*S*'>: *S*,*S*' $\subseteq$ *U* and |S $\cap$ S| = 4}
('four' denotes pairs of sets *S* and *S*' such that *S* and *S*' are subsets of the domain *U* and the cardinality of the intersection of *S* and *S*' is exactly four)

There are two kinds of problems with this reasoning, however. First, it does not follow that just because GQT analyzes 'four' in (1a) as a determiner having a meaning like that in (17), the *lexical* meaning of 'four' must be as specified in (17). Secondly, even if 'four' in (1a) – or indeed *every* occurrence of 'four' – had the meaning suggested in (17), it would not follow that the semantic evidence best supports nominalism.

As for the first problem, it turns out that there are good, independent reasons for thinking that the GQT analysis in (17) is independently flawed. Consider the following example from Krifka (1999), which is ambiguous between at least a *distributive interpretation* given in (46b), and *cumulative interpretation* given in (46c):

(46) a. Three boys ate seven apples.
b. Three boys each ate seven apples, so that twenty one total apples were eaten.
c. Three boys together ate seven apples, so that seven total apples were eaten.

The problem, as Krifka explains, is that because (17) only predicts distributive interpretations, it cannot capture the cumulative interpretation. This suggests that even if GQT is the predominant analysis for natural language quantificational determiners, this provides no compelling reason for thinking that we should adopt that same analysis for 'four' in (1a), let alone (1b).

In fact, in order to explain how cumulative interpretations are possible, Krifka argues that "attributive" uses of 'four' must be understood as *adjectives* expressing cardinal properties of sums of countable individuals, or "atoms" in the sense of Link (1983). On this analysis, "attributive" 'four' in (1a) has something like the meaning suggested in (47), where '#' is a cardinality function mapping sums to numbers representing their atomic parts.

(47) $\lambda P.\lambda x.\ \#(x) = 4 \wedge P(x)$

Ultimately, this affords the following analysis of (46a):[18]

(48) $\exists x.\exists y.\ \#(x) = 3 \wedge \text{boys}(x) \wedge \#(y) = 7 \wedge \text{apples}(y) \wedge \text{ate}(x,y)$

The cumulative interpretation paraphrased in (46c) then arises if the predicate is interpreted *collectively*, so that three boys *together* ate seven apples.

To be clear, the claim is not that "attributive" uses of number expressions must be adjectives rather than determiners because no version of the GQT analysis could, in principle, capture cumulative interpretations. Rather, the semantic case for "attributive" 'four'being an adjective is far more comprehensive in scope. Specifically, as many have noted, 'four' has *many* interrelated uses apart from the "attributive" use witnessed in (1a), including e.g. those in (49).

(49) a. Jupiter's moons are four (in number).
b. No four moons of Jupiter orbit Saturn.

Thus, a desideratum on any empirically adequate semantics for number expressions is that it should not only provide meanings appropriate for all of these uses, but also explain how those meanings are *related*.[19]

Thus, the widespread assumption within linguistic semantics is that number expressions are *polymorphic*, taking on different semantic types in different syntactic environments, thanks to type-shifting (see e.g. Partee (1986a), Landman (2003, 2004), Geurts (2006), Scontras (2014), Kennedy (2015), Rothstein (2013, 2017), and Snyder (2017)). What's more, on all such analyses, 'four' in (1a) and (52a,b) is an *adjective*, and for good reason. On its face, 'four' in (49a) is a predicate, a seemingly appropriate meaning for which is given in (50).

(50) $[[\text{four}]] = \lambda x.\ \#(x) = 4$

The meaning suggested in (47) – appropriate for (49b) – is then derivable from (50) via an independently motivated type-shifting principle, as are meanings potentially appropriate for (1a) and (1b).[20] Crucially, however, determiners *cannot* function as predicates or modifiers – cf. Sect. 3.4.1. This would be entirely mysterious if the

[18] This presupposes type-shifting. See e.g. Rothstein (2017), and Snyder (2017).

[19] Cf. Geurts (2006) and Rothstein (2013).

[20] See Snyder (2017).

lexical meaning of 'four' were that of a determiner since, in that case, the type-shifting principles responsible for generating meanings appropriate for 'four' in (49a,b) would likewise generate meanings appropriate for *all* determiners. In other words, this would incorrectly predict that all determiners can in fact function as predicates and modifiers, contrary to fact.

In short, the problem is not that the GQT analysis provides the wrong *meaning* for 'four' in (1a) – in fact, a meaning equivalent to (17) can be generated from (47) or (50) via commonly accepted type-shifting principles. Rather, the problem is with the inference potentially drawn based on (17): because GQT analyses 'four' in (1a) as denoting a relation between sets, *lexically* 'four' must be a determiner having that same meaning. This is a non-sequitur, as should now hopefully be clear. Yet without some such assumption in place, it simply does not follow that 'four' in (1b) must also have this meaning, even if, to repeat the quote from Hofweber (2005, p. 211) again, "the word 'four' is the same in [(1a)] and [(1a)]".

All of this points towards two important observations relevant to Hofweber's nominalist program. First and foremost, contrary to what Hofweber apparently assumes, it is *not* uncontroversial that "attributive" uses of number expressions, such as 'four' in (1a), are quantificational determiners to be analyzed on the model of GQT. Recall the quote from Hofweber (2007, p. 3–4):

> In contemporary natural-language semantics the uses of 'four' as in [(1a)] are pretty well understood, and 'four' is usually considered to be a determiner, an expression of the same kind as 'some', 'many', and 'all'.

A similar sentiment is expressed in Hofweber (2016, p. 123):

> As it turns out, [GQT] works perfectly well, at least for the cases we are considering here, and it is widely accepted.

It is not clear on what empirical grounds Hofweber could justifiably make either of these assertions. In fact, the first, also endorsed in Hofweber (2016), seemingly belies an understanding of the current state of research within contemporary linguistic theory: the best, most current available evidence points towards 'many' being an *adjective*, not a determiner (see e.g. Rett (2008), Solt (2009), Wellwood (2018), and Snyder (2020)).[21] Again, just because 'many' was analyzed as denoting a relation between sets in Barwise and Cooper (1981), it does not follow that it must be a determiner having that lexical meaning.

More to the point, as the citations above indicate, GQT was not the only analysis of number expressions available at the time of publishing Hofweber (2005), and there was already ample evidence available suggesting that number expressions are better understood as adjectives. Furthermore, while there has been a growing consensus among linguists towards that conclusion ever since, virtually all of this research presupposes, contra Hofweber, that number expressions can function

[21] For one thing, unlike all prototypical determiners, 'many' has a comparative and superlative form – 'more' and 'most', respectively – and is gradable – cf. 'very/so/how many'.

referentially.[22] Thus, it would appear that Hofweber's pronouncements regarding the current state of research within linguistic semantics are at best misleading.

A second significant fact about the polymorphic analyses mentioned above is that they presuppose an independent domain of numbers, to serve as the range of the cardinality function '#'. Specifically, '#' is a *measure function*, or function from entities to numbers. What's more, this features in all meanings relevant to cardinal uses of 'four', including (1a) (cf. (48)). The implication is that even if "attributive" uses do not overtly reference numbers, the metalanguage in which the semantics is formulated is clearly committed to their existence.

However, the same can be said for the GQT analysis. Specifically, (17) contains a numeral ('4'), the referent of which assumed to be a number. In fact, GQT makes rampant use of such numbers to provide a unified analysis of determiner meanings. These include e.g. 'at least four', 'between four and six', and 'four out of five', which explicitly involve number expressions, as well as e.g. 'many', 'most', and 'infinitely many', which do not.[23] Thus, even if 'four' in (1a) had the non-referential meaning attributed in (17), since the metatheory of GQT is committed to numbers, any theorist evoking GQT is also committed to their existence.[24] What's more, those numbers are not obviously eliminable in favor of something else, e.g. the language of first-order quantificational logic. After all, one of the original motivations for GQT was to provide a *uniform* analysis of quantificational expressions, including those which are known to be unanalyzable in terms of first-order quantificational logic, e.g. 'infinitely many'.

In short, even if the lexical meaning of 'four' were that provided in (17), and even if this was the meaning witnessed in (1a,b), it still would not follow that our ordinary number-talk does not involve a commitment to numbers, at least at the metasemantic level. More generally, even if *all* occurrences of number expressions, including their apparent use as numerals in arithmetic statements, were quantificational determiners to be analyzed on the model of GQT, it would not follow that making semantic sense of number talk more generally supports nominalism.

[22] In fact, the only counterexample we are aware of, which happens to be directly informed by, and formulated partially in response to, Hofweber (2005), is Ionin and Matushansky (2006). Incidentally, this also happens to be the target of the semantic arguments mentioned in Sect. 3.4.3.

[23] See Barwise and Cooper (1981).

[24] An anonymous reviewer observes that the same argument would extend to sets, which should be just as objectionable from a nominalist perspective, but that this kind of commitment might be avoided by appealing to a pluralist metalanguage, perhaps following Boolos (1985). As far as we know, whether all of GQT can be recovered within a pluralist metalanguage is an open question, though McKay (2006) makes progress in this direction. Even so, the question would remain as to whether an empirically adequate, nominalist-friendly pluralist semantics for number expressions could be formulated, something which some of us have cast doubt on in other work (e.g Snyder and Shapiro 2021).

3.5 Conclusion

We have argued that neither linguistic component of Hofweber's analysis of ordinary number talk survives empirical scrutiny. Specifically, the syntactic component fails in virtue of the empirical implausibility of the operation posited ("extraction"), while many of the distinctive semantic theses put forward by Hofweber are empirically problematic. These include:

(ST1) 'four' in (1a) is a non-referential determiner, to be analyzed on the model of GQT.
(ST2) 'four' in (1b) has the same non-referential meaning witnessed in (1a).
(ST3) Numerals, or at least those occurring in arithmetic statements, are semantically bare determiners.

Some of these problems arguably stem from an initial syntactic misclassification encoded in (ST1), namely that 'four' in Frege's (1a) is a determiner, rather than an adjective. Without that initial assumption in place, (ST2) clearly doesn't follow, even if we grant "extraction". On the other hand, we have seen that it is potentially important for Hofweber's larger program that 'four' in (1a) be seen as a determiner, as at least certain adjectives appear to have genuinely referential uses, unlike all known determiners.

Furthermore, in addition to the numerous problems noted for analyzing numerals in arithmetic statements as semantically bare determiners, the empirical motivation for (ST3) is further weakened once we recognize that a variety of expressions can be coordinated in a manner seemingly resembling (51).

(51) Three and two is five.

Consider the examples in (52), for instance, respectively involving color expressions, measure phrases, and bare nouns.

(52) a. Red and blue is purple.
b. Two feet and twelve inches is one foot.
c. Horseradish and ketchup is cocktail sauce.

Naively, all four examples have a distinctly "combinatory" feel: the result of combining the pre-copular things results in the post-copular thing. Seen this way, nothing about (51) itself forces the conclusion that the number expressions involved are (semantically bare) determiners, and thus non-referential expressions. In fact, the expressions in (52a–c) are commonly assumed within linguistic semantics to have genuinely referential uses.[25] What's more, it has been argued, notably by Rothstein (2013, 2017), that the same semantic operation responsible for the referentiality of the expressions in (52) – *nominalization* – is also responsible for the referentiality

[25] See e.g. Scontras (2014) for measure phrases, and Chierchia (1998) for bare nouns.

of numerals. Thus, given that a *uniform*, compositional analysis of (51) and (52a-c) is independently desirable, providing an empirically adequate semantics for (51) might well require that the apparent numerals involved are genuine singular terms.[26]

All of this casts significant doubt on the empirical motivations for Hofweber's Adjectivalism, and with it the proposed resolutions of Frege's Other Puzzle and the Easy Argument. Ultimately, this highlights the difficulties inherent in the sort of empirically informed methodological naturalism that Hofweber's project intends to engage in.

In our view, Hofweber's analysis is thus perhaps best viewed as an impressive exploration of an intriguing linguistic hypothesis that, if true, could have significant ontological consequences for the philosophy of mathematics. Specifically, if all uses of number expressions could be viewed as non-referential determiners, then making semantic sense of number talk more generally might not require an ontology of natural numbers. It's just that, given the best available linguistic evidence, the antecedent of this conditional is highly implausible. The takeaway lesson is that insofar as one seeks to engage in this sort of methodological naturalism, as we intend to do, one must ignore prior metaphysical predilections and let the empirical chips fall where they may.

References

Abels, K. 2017. Displacement in syntax. In *Oxford Research Encyclopedia of Linguistics*, ed. M. Aronoff. New York: Oxford University Press.

Balcerak-Jackson, B. 2013. Defusing easy arguments for numbers. *Linguistics and Philosophy* 36: 447–461.

Barwise, J., and R. Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 1: 413–458.

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

Boolos, G. 1985. Nominalist platonism. *The Philosophical Review 94*: 327–344.

Breheny, R. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25: 93–139.

Brogaard, B. 2007. Number words and ontological commitment. *Philosophical Quarterly* 57: 1–20.

Chierchia, G. 1998. Reference to kinds across language. *Natural Language Semantics 6*: 339–405.

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

Felka, K. 2014. Number words and reference to numbers. *Philosophical Studies* 168: 261–268.

Frege, G. 1884. *Grundlagen der Arithmetik.*

Geurts, B. 2006. Take 'five'. In *Non-Definiteness and Plurality*, ed. S. Vogleer and L. Tasmowski, 311–329. Amsterdam: Benjamins.

Hale, B. 1987. *Abstract Objects*. Oxford: Basil Blackwell.

Higgins, R.F. 1973. *The Pseudo-cleft Construction in English*, Garland.

Hodes, H. 1984. Logicism and the ontological commitments of arithmetic. *The Journal of Philosophy* 81: 123–149.

Hofweber, T. 2005. Number determiners, numbers, and arithmetic. *The Philosophical Review* 114: 179–225.

[26] Contra Moltmann (2013).

———. 2007. Innocent statements and their metaphysically loaded counterparts. *Philosopher's Imprint* 7: 1–33.
———. 2014. Extraction, displacement, and focus. *Linguistics and Philosophy* 37: 263–267.
———. 2016. *Ontology and the Ambitions of Metaphysics*. Oxford University Press, New York.
Horn, L. 1972. *On the Semantic Properties of Logical Operators*. PhD thesis, University of California, Los Angeles.
Ionin, T., and O. Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics 23*: 315–360.
Kennedy, C. 2015. A "de-Fregean" semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics* 10: 1–44.
Kennedy, C., and L. McNally. 2010. Color, context, and compositionality. *Synthese 174*: 79–98.
Kratzer, A., and I. Heim. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.
Krifka, M. 1999. At least some determiners aren't determiners. In *The Semantics/Pragmatics Interface from Different Points of View, 1*, 257–291.
Landman, F. 2003. Predicate-argument mismatches and the adjectival theory of indefinites. In *NP to DP*, ed. M. Coene and Y. D'hulst. Amsterdam: John Benjamins.
———. 2004. *Indefinites and the Type of Sets*. Malden: Blackwell.
Leng, M. 2010. *Mathematics and Reality*. Oxford: Oxford University Press.
Link, G. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Formal Semantics: The Essential Readings*, ed. P. Portner and B. Partee, 127–146. Oxford: Blackwell.
McKay, T. 2006. *Plural Predication*. New York: Oxford University Press.
McNally, L. 2011. Color terms: A case study in natural language ontology. In *Workshop on the Syntax and Semantics of Nounhood and Adjectivehood.* Barcelona.
McNally, L., and H. de Swart. 2011. Inflection and derivation: How adjectives and nouns refer to abstract objects. In *Proceedings of the 18th Amsterdam Colloquium*, 425–434.
Mikkelsen, L. 2005. *Copular Clauses: Specification, Predication, and Equation*. Amsterdam: Benjamins.
Moltmann, F. 2013. Reference to numbers in natural language. *Philosophical Studies* 162: 499–536.
Partee, B. 1986a. Noun phrase interpretation and type-shifting principles. In *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, ed. J. Groenendijk, D. de Jongh, and M. Stokhof. Dordrecht: Foris.
———. 1986b. Ambiguous pseudoclefts with unambiguous be. In *Proceedings of NELS 16*, ed. S. Berman, J. Choe, and J. McDonough, 354–366. Amherst: GLSA.
Rett, J. 2008. *Degree Modification in Natural Language*, PhD dissertation, Rutgers University.
Romero, M. 2005. Concealed questions and specificational subjects. *Linguistics and Philosophy* 28: 687–737.
Rothstein, S. 2013. A Fregean semantics for number words. In *Proceedings of the 19th Amsterdam Colloquium*, 179–186. Amsterdam: Universiteit van Amsterdam.
———. 2017. *Semantics for Counting and Measuring*. Cambridge: Cambridge University Press.
Schlenker, P. 2003. Clausal equations. *Natural Language and Linguistic Theory* 21: 157–214.
Scontras, G. 2014. *The semantics of measurement*. Doctoral dissertation, Harvard University.
Snyder, E. 2017. Numbers and cardinalities: What's really wrong with the easy argument? *Linguistics and Philosophy* 40: 373–400.
———. 2020. Counting, measuring, and the fractional cardinalities puzzle. *Linguistics and Philosophy.* 44: 513.
Snyder, Eric, Richard Samuels, and Stewart Shapiro 2021. Resolving Frege's other puzzle, *Philosophia Mathematica*, forthcoming.
Snyder, Eric and Stewart Shapiro. 2021. Groups, sets, and paradox. *Linguistics and Philosophy*, forthcoming.
Solt, S. 2009. *The Semantics of Adjectives of Quantity*, PhD dissertation, City University of New York.

Wellwood, A. 2018. Structure preservation in comparatives. *Semantics and Linguistic Theory* 28: 78–99.

Wright, C. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

Yablo, S. 2005. The myth of the seven. In *Fictionalist Approaches to Metaphysics*, ed. M. Kalderon, 90–115. Oxford: Oxford University Press.

Chapter 4
Exploring Mathematical Objects from Custom-Tailored Mathematical Universes

Ingo Blechschmidt

Abstract Toposes can be pictured as mathematical universes. Besides the standard topos, in which most of mathematics unfolds, there is a colorful host of alternate toposes in which mathematics plays out slightly differently. For instance, there are toposes in which the axiom of choice and the intermediate value theorem from undergraduate calculus fail. The purpose of this contribution is to give a glimpse of the toposophic landscape, presenting several specific toposes and exploring their peculiar properties, and to explicate how toposes provide distinct lenses through which the usual mathematical objects of the standard topos can be viewed.

Keywords Topos theory · Realism debate · Well-adapted language · Constructive mathematics

Toposes can be pictured as mathematical universes in which we can do mathematics. Most mathematicians spend all their professional life in just a single topos, the so-called *standard topos*. However, besides the standard topos, there is a colorful host of alternate toposes which are just as worthy of mathematical study and in which mathematics plays out slightly differently (Fig. 4.1).

For instance, there are toposes in which the axiom of choice and the intermediate value theorem from undergraduate calculus fail, toposes in which any function $\mathbb{R} \to \mathbb{R}$ is continuous and toposes in which infinitesimal numbers exist.

The purpose of this contribution is twofold.

1. We give a glimpse of the toposophic landscape, presenting several specific toposes and exploring their peculiar properties.
2. We explicate how toposes provide distinct lenses through which the usual mathematical objects of the standard topos can be viewed.

I. Blechschmidt (✉)
Institut für Mathematik, Universität Augsburg, Augsburg, Germany
e-mail: ingo.blechschmidt@math.uni-augsburg.de

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_4

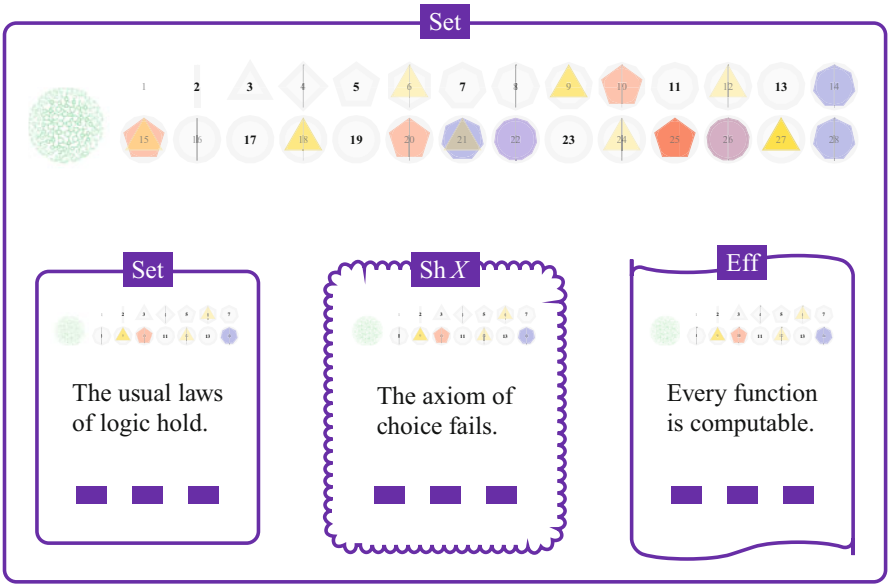


Fig. 4.1 A glimpse of the toposophic landscape, displaying alongside the standard topos Set two further toposes

Viewed through such a lens, a given mathematical object can have different properties than when viewed normally. In particular, it can have better properties for the purposes of specific applications, especially if the topos is custom-tailored to the object in question. This change of perspective has been used in mathematical practice. To give just a taste of what is possible, through the lens provided by an appropriate topos, any given ring can look like a field and hence mathematical techniques for fields also apply, through the lens, to rings.

We argue that toposes and specifically the change in perspective provided by toposes are ripe for philosophical analysis. In particular, there are the following connections with topics in the philosophy of mathematics:

1. Toposes enrich the realism/anti-realism debate in that they paint the larger picture that the platonic heaven of mathematical objects is not unique: besides the standard heaven of the standard topos, we can fathom the alternate heavens of all other toposes, all embedded in a second-order heaven.
2. To some extent, the mathematical landscape depends on the commonly agreed-upon rules of mathematics. These are not entirely absolute; for instance, it is conceivable that from the foundational crisis Brouwer's intuitionism would have emerged as the main school of thought and that we would now all reject the law of excluded middle. Toposes allow us to explore alternatives to how history has played out.

3. Mathematics is not only about studying mathematical objects, but also about studying the relations between mathematical objects. The distinct view on mathematical objects provided by any topos uncovers relations which otherwise remain hidden.[1]
4. In some cases, a mathematical relation can be expressed quite succinctly using the language of a specific topos and not so succinctly using the language of the standard topos. This phenomenon showcases the importance of *appropriate language*.
5. Toposes provide new impetus to study constructive mathematics and intuitionistic logic, in particular also to restrict to intuitionistic logic on the meta level and to consider the idea that the platonic heaven might be governed by intuitionistic logic.

   We invite further research on these connections.

We intend this contribution to be self-contained and do not assume familiarity with topos theory or category theory, having a diverse readership of people interested in philosophy of mathematics in mind. However, to make this text more substantial to categorically-inclined readers, some categorical definitions are included. These definitions can be skipped without impacting the main message of this contribution.

Readers who would like to learn more details are directed to the survey of category theory by Marquis (2019) and to a gentle introduction to topos theory by Leinster (2011). Standard references for the internal language of toposes include Mac Lane and Moerdijk (1992, Chapter VI), Goldblatt (1984, Chapter 14), Caramello (2014), Streicher (2004), Shulman (2016), Borceux (1994, Chapter 6) and Johnstone (2002, Part D).

Other aspects of toposes This note focuses on just a single aspect of toposes, the view of toposes as alternate mathematical universes. This aspect is not the only one, nor did it historically come first.

Toposes were originally conceived by Grothendieck in the early 1960s for the needs of algebraic geometry, as a general framework for constructing and studying invariants in classical and new geometric contexts, and it is in that subject that toposes saw their deepest applications. The proof of Fermat's Last Theorem is probably the most prominent such application, crucially resting on the cohomology and homotopy invariants provided by toposes.

In the seminal work introducing toposes by Artin et al. (1972), toposes are viewed as generalized kinds of spaces. Every topological space X gives rise to a topos, the *topos of sheaves over* X, and every continuous map gives rise to a *geometric morphism* between the induced sheaf toposes, but not every topos is of this form. While the open sets of a topological space are required to be parts of the

[1] The research program put forward by Caramello (2018) provides a further topos-theoretic way for uncovering hidden relations, though not between objects but between mathematical theories (Chap. 9).

space, the opens of toposes are not; and while for open subsets U and V there is only a truth value as to whether U is contained in V, in a general topos there can be many distinct ways how an open is contained in another one. This additional flexibility is required in situations where honest open subsets are rare, such as when studying the étale cohomology of a scheme as in Milne (2013).

That toposes could also be regarded as mathematical universes was realized only later, by Bill Lawvere and Myles Tierney at the end of the 1960s. They abstracted some of the most important categorical properties of Grothendieck's toposes into what is now known as the definition of an *elementary topos*. Elementary toposes are considerably more general and less tied to geometry than the original toposes. The theory of elementary toposes has a substantially different, logical flavor, not least because a different notion of morphism plays an important role. To help disambiguate, there is a trend to rename elementary toposes to *logoses*, but this text still follows the standard convention.

A further perspective on toposes emerged in the early 1970s with the discovery that toposes can be regarded as embodiments of a certain kind of first-order theories, the *geometric theories* briefly discussed on page 71. The so-called *classifying toposes* link geometrical and logical aspects and are fundamental to Olivia Caramello's bridge-building program set out in Caramello (2018). Geometrically, the classifying topos of a geometric theory $\mathbb{T}$ can be regarded as the generalized space of models of $\mathbb{T}$; this idea is due to Hakim (1972), though she did not cast her discovery in this language. Logically, the classifying topos of $\mathbb{T}$ can be regarded as a particular mathematical universe containing the *generic* $\mathbb{T}$*-model*, a model which has exactly those properties which are shared by all models.

Yet more views on toposes are fruitfully employed – Johnstone (2002, pages vii–viii) lists ten more – but we shall not review them here. A historical survey was compiled by McLarty (1990).

4.1 Toposes as Alternate Mathematical Universes

A topos is a certain kind of *category*, containing objects and morphisms between those objects. The precise definition is recorded here only for reference. Appreciating it requires some amount of category theory, but, as will be demonstrated in the following sections, exploring the mathematical universe of a given topos does not.

Definition 1 A *topos* is a category which has all finite limits, is cartesian closed, has a subobject classifier and contains a natural numbers object.[2]

Put briefly, these axioms state that a topos should share several categorical properties with the category of sets; they ensure that each topos contains its own

[2] More precisely, this is the definition of an *elementary topos with a natural numbers object*. Since this definition is less tied to geometry than Grothendieck's (as categories of sheaves over sites), there is a trend to call these toposes *logoses*. However, that term also has other uses.

versions of familiar mathematical objects such as natural numbers, real numbers, groups and manifolds, and is closed under the usual constructions such as cartesian products or quotients. The prototypical topos is the standard topos:

Definition 2 The *standard topos* Set is the category which has all sets as its objects and all maps between sets as morphisms.

Given a topos $\mathcal{E}$, we write "$\mathcal{E} \models \varphi$" to denote that a mathematical statement φ *holds in* $\mathcal{E}$. The meaning of "$\mathcal{E} \models \varphi$" is defined by recursion on the structure of φ following the so-called *Kripke–Joyal translation rules*. For instance, the rules for translating conjunction and falsity read

$$\begin{aligned} &\mathcal{E} \models (\alpha \wedge \beta) && \text{iff } \mathcal{E} \models \alpha \quad \text{and} \quad \mathcal{E} \models \beta, \\ &\mathcal{E} \models \bot && \text{iff } \mathcal{E} \text{ is the trivial topos.} \end{aligned}$$

The remaining translation rules are more involved, as detailed by Mac Lane and Moerdijk (1992, Section VI.7); we do not list them here for the case of a general topos $\mathcal{E}$, but we will state them in the next sections for several specific toposes. We refer to "$\mathcal{E} \models \varphi$" also as the "external meaning of the internal statement φ".

In the definition of $\mathcal{E} \models \varphi$, the statement φ can be any statement in the language of a general version of higher-order predicate calculus with dependent types, with a base type for each object of $\mathcal{E}$ and with a constant of type X for each morphism $1 \to X$ in $\mathcal{E}$. In practice almost any mathematical statement can be interpreted in a given topos.[3] We refrain from giving a precise definition of the language here, but refer to the references Shulman (2010, Section 7) and Mac Lane and Moerdijk (1992, Section VI.7) for details.

It is by the Kripke–Joyal translation rules that we can access the alternate universe of a topos. In the special case of the standard topos Set, the definition of "Set $\models \varphi$" unfolds to φ for any statement φ. Hence a statement holds in the standard topos if and only if it holds in the usual mathematical sense.

4.1.1 The Logic of Toposes

By their definition as special kinds of categories, toposes are merely algebraic structures not unlike groups or vector spaces. Hence we need to argue why we picture toposes as mathematical universes while we do not elevate other kinds of

[3] The main exceptions are statements from set theory, which typically make substantial use of a global membership predicate "$\in$". Toposes only support a typed *local* membership predicate, where we may write "$x \in A$" only in the context of some fixed type M such that x is of type M and A is of type $P(M)$, the power type of M. We refer to Fourman (1980), Streicher (2009), and Awodey et al. (2014) for ways around this restriction.

algebraic structures in the same way. For us, this usage is justified by the following metatheorem:

Theorem 1 *Let $\mathcal{E}$ be a topos and let φ be a statement such that $\mathcal{E} \models \varphi$. If φ intuitionistically entails a further statement ψ (that is, if it is provable in intuitionistic logic that φ entails ψ), then $\mathcal{E} \models \psi$.*

This metatheorem allows us to *reason* in toposes. When first exploring a new topos $\mathcal{E}$, we need to employ the Kripke–Joyal translation rules each time we want to check whether a statement holds in $\mathcal{E}$. But as soon as we have amassed a stock of statements known to be true in $\mathcal{E}$, we can find more by deducing their logical consequences.

For instance, in any topos where the statement "any map $\mathbb{R} \to \mathbb{R}$ is continuous" is true, also the statement "any map $\mathbb{R} \to \mathbb{R}^2$ is continuous" is, since there is an intuitionistic proof that a map into a higher-dimensional Euclidean space is continuous if its individual components are.

The only caveat of Theorem 1 is that toposes generally only support intuitionistic reasoning and not the full power of the ordinary *classical reasoning*. That is, within most toposes, the law of excluded middle ($\varphi \vee \neg\varphi$) and the law of double negation elimination ($\neg\neg\varphi \Rightarrow \varphi$) are not available. It is intuitionistic logic and not classical logic which is the common denominator of all toposes; we cannot generally argue by contradiction in a topos.

While it may appear that these two laws pervade any mathematical theory, in fact a substantial amount of mathematics can be developed intuitionistically (see for instance Mines et al. (1988) and Lombardi and Quitté (2015) for constructive algebra, Bishop and Bridges (1985) for constructive analysis and Bauer (2012), Bauer (2013), and Melikhov (2015) for accessible surveys on appreciating intuitionistic logic) and hence the alternate universes provided by toposes cannot be too strange: In any topos, there are infinitely many prime numbers, the square root of two is not rational, the fundamental theorem of Galois theory holds and the powerset of the naturals is uncountable.

That said, intuitionistic logic still allows for a considerable amount of freedom, and in many toposes statements are true which are baffling if one has only received training in mathematics based on classical logic. For instance, on first sight it looks like the sign function

$$\operatorname{sgn} : \mathbb{R} \longrightarrow \mathbb{R},\ x \longmapsto \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{if } x > 0, \end{cases}$$

is an obvious counterexample to the statement "any map $\mathbb{R} \to \mathbb{R}$ is continuous". However, a closer inspection reveals that the sign function cannot be proven to be a total function $\mathbb{R} \to \mathbb{R}$ if only intuitionistic logic is available. The domain of the sign function is the subset $\{x \in \mathbb{R} \mid x < 0 \vee x = 0 \vee x > 0\} \subseteq \mathbb{R}$, and in intuitionistic logic this subset cannot be shown to coincide with $\mathbb{R}$.

Sections 4.2, 4.3, and 4.4 present several examples for such anti-classical statements and explain how to make sense of them. There are also toposes which are closer to the standard topos and do not validate such anti-classical statements:

Definition 3 A topos $\mathcal{E}$ is *boolean* if and only if the laws of classical logic are true in $\mathcal{E}$.

Since exactly those statements hold in the standard topos which hold on the meta level, the standard topos is boolean if and only if, as is commonly supposed, the laws of classical logic hold on the meta level. Most toposes of interest are not boolean, irrespective of one's philosophical commitments about the meta level, and conversely some toposes are boolean even if classical logic is not available on the meta level.

Remark 1 The axiom of choice (which is strictly speaking not part of classical logic, but of classical set theory) is also not available in most toposes. By *Diaconescu's theorem*, the axiom of choice implies the law of excluded middle in presence of other axioms which are available in any topos.

At this point in the text, all prerequisites for exploring toposes have been introduced. The reader who wishes to develop, by explicit examples, intuition for working internally to toposes is invited to skip ahead to Sect. 4.2.

4.1.2 Relation to Models of Set Theory

In set theory, philosophy and logic, models of set theories are studied. These are structures $(M, \in)$ validating the axioms of some set theory such as Zermelo–Fraenkel set theory with choice ZFC, and they can be pictured as "universes in which we can do mathematics" in much the same way as toposes.

In fact, to any model $(M, \in)$ of a set theory such as ZF or ZFC, there is a topos Set_M such that a statement holds in Set_M if and only if it holds in M.[4]

Example 1 The topos Set_V associated to the universe V of all sets (if this structure is available in one's chosen ontology) coincides with the standard topos Set.

In set theory, we use forcing and other techniques to construct new models of set theory from given ones, thereby exploring the set-theoretic multiverse. There are similar techniques available for constructing new toposes from given ones, and some of these correspond to the techniques from set theory.

[4] The topos Set_M can be described as follows: Its objects are the elements of M, that is the entities which M believes to be sets, and its morphisms are those entities which M believes to be maps. The topos Set_M validates the axioms of the structural set theory ETCS, see McLarty (2004), Marquis (2013), and Barton and Friedman (2019), and models are isomorphic if and only if their associated toposes are equivalent as categories, see Mac Lane and Moerdijk (1992, Section VI.10).

However, there are also important differences between the notion of mathematical universes as provided by toposes and as provided by models of set theory, both regarding the subject matter and the reasons for why we are interested in them.

Firstly, toposes are more general than models of set theory. Every model of set theory gives rise to a topos, but not every topos is induced in this way from a model of set theory. Unlike models of ZFC, most toposes do not validate the law of excluded middle, much less so the axiom of choice.

Secondly, there is a shift in emphasis. An important philosophical objective for studying models of set theory is to explore which notions of sets are coherent: Does the cardinality of the reals need to be the cardinal directly succeeding $\aleph_0$, the cardinality of the naturals? No, there are models of set theory in which the continuum hypothesis fails. Do non-measurable sets of reals need to exist? No, in models of ZF+AD, Zermelo–Fraenkel set theory plus the axiom of determinacy, it is a theorem that every subset of $\mathbb{R}^n$ is Lebesgue-measurable. Can the axiom of choice be added to the axioms of ZF without causing inconsistency? Yes, if M is a model of ZF then L^M, the structure of the constructible sets of M, forms a model of ZFC.

Toposes can be used for similar such purposes, and indeed have been, especially to explore the various intuitionistic notions of sets. However, an important aspect of topos theory is that toposes are used to explore the *standard* mathematical universe: truth in the effective topos tells us what is computable; truth in sheaf toposes tells us what is true locally; toposes adapted to synthetic differential geometry can be used to rigorously work with infinitesimals. All of these examples will be presented in more detail in the next sections.

In a sense which can be made precise, toposes allow us to study the usual objects of mathematics from a different point of view – one such view for every topos – and it is a beautiful and intriguing fact that with the sole exception of the law of excluded middle, the laws of logic apply to mathematical objects also when viewed through the lens of a specific topos.

4.1.3 A Glimpse of the Toposophic Landscape

There is a proper class of toposes. Figure 4.1 depicts three toposes side by side: the standard topos, a sheaf topos and the effective topos. Each of these toposes tells a different story of mathematics, and any topos which is not the standard topos invites us to ponder alternative ways how mathematics could unfold.

Some of the most prominent toposes are the following.

1. The *trivial topos*. In the trivial topos, any statement whatsoever is true. The trivial topos is not interesting on its own, but its existence streamlines the theory and it can be an interesting question whether a given topos coincides with the trivial topos.
2. Set, the *standard topos*. A statement is true in Set iff it is true in the ordinary mathematical sense.

3. Set_M, the topos associated to any model $(M, \in)$ of ZF.
4. Set^W, the category of functors $(W, \leq) \to$ Set associated to any Kripke model $(W, \leq)$. A statement is true in this topos iff it is valid with respect to the ordinary Kripke semantics of $(W, \leq)$. This example shows that the Kripke–Joyal semantics of toposes generalizes the more familiar Kripke semantics.
5. Eff, the *effective topos*. A statement is true in Eff iff if it has a *computable witness* as detailed in Sect. 4.2. In Eff, any function $\mathbb{N} \to \mathbb{N}$ is computable, any function $\mathbb{R} \to \mathbb{R}$ is continuous and the countable axiom of choice holds (even if it does not on the meta level).
6. $\mathrm{Sh}(X)$, the *topos of sheaves* over any space X. A statement is true in $\mathrm{Sh}(X)$ iff it holds *locally on* X, as detailed in Sect. 4.3. For most choices of X, the axiom of choice and the intermediate value theorem fail in $\mathrm{Sh}(X)$, and this failure is for geometric reasons.
7. $\mathrm{Zar}(A)$, the *Zariski topos* of a ring A presented in Sect. 4.4. This topos contains a mirror image of A which is a field, even if A is not.
8. $\mathrm{Bohr}(A)$, the *Bohr topos* associated to a noncommutative C^*-algebra A. This topos contains a mirror image of A which is commutative. In this sense, quantum mechanical systems (which are described by noncommutative C^*-algebras) can be regarded as classical mechanical systems (which are described by commutative algebras). Details are described by Butterfield et al. (1998) and Heunen et al. (2009).
9. $\mathrm{Set}[\mathbb{T}]$, the *classifying topos* of a geometric theory $\mathbb{T}$.[5] This topos contains the *generic* $\mathbb{T}$*-model*. For instance, the classifying topos of the theory of groups contains the *generic group*. Arguably it is this group which we implicitly refer to when we utter the phrase "Let G be a group.". The generic group has exactly those properties which are shared by any group whatsoever.[6]
10. $T(\mathcal{L}_0)$, the *free topos*. A statement is true in the free topos iff it is intuitionistically provable. Lambek and Scott proposed that the free topos can reconcile moderate platonism (because this topos has a certain universal property which can be used to single it out among the plenitude of toposes), moderate formalism (because it is constructed in a purely syntactic way) and moderate logicism (because, as a topos, it supports an intuitionistic type theory). Details are described by Lambek (1994) and Couture and Lambek (1991).

There are several constructions which produce new toposes from a given topos $\mathcal{E}$. A non-exhaustive list is the following.

[5] A geometric theory is a theory in many-sorted first-order logic whose axioms can be put as *geometric sequents*, sequents of the form $\varphi \vdash_{\mathbf{x}} \psi$ where φ and ψ are geometric formulas (formulas built from equality and specified relation symbols by the logical connectives $\top \bot \wedge \vee \exists$ and by arbitrary set-indexed disjunctions $\bigvee$).

[6] More precisely, this is only true for those properties which can be formulated as geometric sequents. For arbitrary properties φ, the statements "the generic group has property φ" and "all groups have property φ" need not be equivalent. This imbalance has mathematical applications and is explored in Blechschmidt (2020).

1. Given an object X of $\mathcal{E}$, the *slice topos* $\mathcal{E}/X$ contains the *generic element* x_0 of X. This generic element can be pictured as the element we implicitly refer to when we utter the phrase "Let x be an element of X.". A statement $\varphi(x_0)$ about x_0 is true in $\mathcal{E}/X$ if and only if in $\mathcal{E}$ the statement $\forall x : X.\ \varphi(x)$ is true.
   For instance, the topos $\mathrm{Set}/\mathbb{Q}$ contains the generic rational number x_0. Neither the statement "x_0 is zero" nor the statement "x_0 is not zero" hold in $\mathrm{Set}/\mathbb{Q}$, as it is neither the case that any rational number in Set is zero nor that any rational number in Set is not zero. Like any rational number, the number x_0 can be written as a fraction $\frac{a}{b}$. Just as x_0 itself, the numbers a and b are quite indetermined.
2. Given a statement φ (which may contain objects of $\mathcal{E}$ as parameters but which must be formalizable as a geometric sequent), there is a largest subtopos of $\mathcal{E}$ in which φ holds. This construction is useful if neither φ nor $\neg\varphi$ hold in $\mathcal{E}$ and we want to force φ to be true. If $\mathcal{E} \models \neg\varphi$, then the resulting topos is the trivial topos. (A subtopos is not simply a subcategory; rather, it is more like a certain kind of quotient category. We do not give, and for the purposes of this contribution do not need, further details.)
3. There is a "smallest dense" subtopos $\mathrm{Sh}_{\neg\neg}(\mathcal{E})$. This topos is always boolean, even if $\mathcal{E}$ and the meta level are not. For a mathematician who employs intuitionistic logic on their meta level, the nonconstructive results of their classical colleagues do not appear to make sense in Set, but they hold in $\mathrm{Sh}_{\neg\neg}(\mathrm{Set})$. If classical logic holds on the meta level, then Set and $\mathrm{Sh}_{\neg\neg}(\mathrm{Set})$ coincide.
   The topos $\mathrm{Sh}_{\neg\neg}(\mathcal{E})$ is related to the *double negation translation* from classical logic into intuitionistic logic: A statement holds in $\mathrm{Sh}_{\neg\neg}(\mathcal{E})$ if and only if its translation holds in E, see Blechschmidt (2017, Theorem 6.31).

Toposes are still mathematical structures, and as long as we study toposes within the usual setup of mathematics, our toposes are all part of the standard topos. This is why Fig. 4.1 pictures the standard topos twice, once as a particular topos next to others, and once as the universe covering the entirety of our mathematical discourse.[7] The toposes which we can study in mathematics do not tell us all possible stories how mathematics could unfold, only those which appear coherent from the point of view of the standard topos, and the topos-theoretic multiverse which we have access to is just a small part of an even larger landscape.[8]

[7] There is a fine print to consider. Technically, if we work within ZF or its intuitionistic cousin IZF, most toposes of interest are proper classes, not sets. In particular Set itself is a proper class. Hence Fig. 4.1 should not be interpreted as indicating that toposes are contained in Set as objects, which most are not. In this regard toposes are similar to class-sized inner models in set theory. We believe that the vague statement "our toposes are all part of the standard topos" is still an apt description of the situation. A possible formalization is (the trivial observation that) "our toposes are all indexed categories over Set".

[8] This paragraph employs an overly narrow conception of "mathematics", focusing only on those mathematical worlds which form toposes and for instance excluding any predicative flavors of mathematics (Laura Crosilla's survey in Crosilla (2018) is an excellent introduction). Toposes are impredicative in the sense that any object of a topos is required to have a powerobject. A predicative cousin of toposes are the *arithmetic universes* introduced by Joyal which have recently been an

To obtain just a hint of how the true landscape looks like, we can study topos theory from the inside of toposes; the resulting picture can look quite different from the picture which emerges from within the standard topos.

For instance, from within the standard topos, we can write down the construction which yields the standard topos and the construction which yields the effective topos Eff and observe that the resulting toposes are not at all equivalent: In Eff, any function $\mathbb{R} \to \mathbb{R}$ is continuous while Set abounds with discontinuous functions (at least if we assume a classical meta level). In contrast, if we carry out these two constructions from within the effective topos, we obtain toposes which are elementarily equivalent. More precisely, for any statement φ of higher order arithmetic,

$$\text{Eff} \models (\text{Set} \models \varphi) \qquad \text{iff} \qquad \text{Eff} \models (\text{Eff} \models \varphi).$$

In this sense the construction which yields the effective topos is *idempotent* (van Oosten, 2008, Section 3.8.3).

Remark 2 The picture of a topos-theoretic multiverse is related to Hamkin's multiverse view in set theory as put forward in Hamkins (2012). In fact, the topos-theoretic multiverse can be regarded as an extension of the set-theoretic multiverse: While Hamkins proposes to embrace all models of set theory (not necessarily all of them equally – we might prefer some models over others), we propose to embrace all toposes (again not necessarily all of them equally). As every model M of set theory gives rise to a topos Set_M, the set-theoretic multiverse is contained in the topos-theoretic one.

However, a central and intriguing feature of the multiverse view in set theory has, as of yet, no counterpart in topos theory: namely a systematic study of its modal logic with respect to various notions of relations between toposes.

4.1.4 A Syntactic Account of Toposes

We introduced toposes from a semantic point of view. There is also a second, purely syntactic point of view on toposes:

1. (semantic view) A topos is an alternate mathematical universe. Any topos contains its own stock of mathematical objects. A "transfer theorem" relates properties of those objects with properties of objects of the standard topos: A statement φ about the objects of a topos $\mathcal{E}$ holds in $\mathcal{E}$ iff the statement "$\mathcal{E} \models \varphi$" holds in the standard topos.

---

important object of consideration by Maietti and Vickers, see Maietti (2010), Maietti and Vickers (2012), and Vickers (2016).

2. (syntactic view) A topos is merely an index to a syntactical translation procedure. Any topos $\mathcal{E}$ gives rise to a "generalized modal operator" which turns a statement φ (about ordinary mathematical objects) into the statement "$\mathcal{E} \models \varphi$" of the same kind (again about ordinary mathematical objects).

For instance, in the semantic view, the effective topos is an alternative universe which contains its own version of the natural numbers. These naturals cannot be directly compared with the naturals of the standard topos, for they live in distinct universes, but by the transfer theorem they are still linked in a nontrivial way: For instance, the statement "there are infinitely many primes in Eff" (a statement about natural numbers in Eff) is equivalent to the statement "for any number n, there *effectively* exists a prime number $p > n$" (a statement about natural numbers and computability in the standard topos). (The meaning of *effectivity* will be recalled in Sect. 4.2.)

In the syntactic view, the effective topos merely provides a coherent way of adding the qualification "effective" to mathematical statements, for instance turning the statement "for any number n, there exists a prime number $p > n$" into "for any number n, there *effectively* exists a prime number $p > n$". Similarly, a sheaf topos $\mathrm{Sh}(X)$ provides a coherent way for turning statements about real numbers and real functions into statements about continuous X-indexed families of real numbers and real functions.

The crucial point is that the translation scheme provided by any topos is sound with respect to intuitionistic logic. Hence, regardless of our actual position on toposes as alternate universes, working under the lens of a given topos *feels like* working in an alternate universe.

4.2 The Effective Topos, a Universe Shaped by Computability

A basic question in computability theory is: Which computational tasks are solvable in principle by computer programs? For instance, there is an algorithm for computing the greatest common divisor of any pair of natural numbers, and hence we say "any pair of natural numbers *effectively* has a greatest common divisor" or "the function $\mathbb{N} \times \mathbb{N} \to \mathbb{N},\ (n, m) \mapsto \gcd(n, m)$ is *computable*".

In such questions of computability, practical issues such as resource constraints or hardware malfunctions are ignored; we employ the theoretical notion of *Turing machines*, a mathematical abstraction of the computers of the real world.

A basic observation in computability theory is that there are computational tasks which are not solvable even for these idealized Turing machines. The premier example is the *halting problem*: Given a Turing machine M, determine whether M terminates (comes to a stop after having carried out finitely many computational steps) or not.

A Turing machine H which would solve this problem, that is read the description of a Turing machine M as input and output one or zero depending on whether M terminates or not, would be called a *halting oracle*, and a basic result is that there are no halting oracles. If we fix some effective enumeration of all Turing machines, then we can express the undecidability of the halting problem also by saying that the *halting function*

$$h : \mathbb{N} \longrightarrow \mathbb{N},\ n \longmapsto \begin{cases} 1, & \text{if the } n\text{-th Turing machine terminates,} \\ 0, & \text{otherwise,} \end{cases}$$

is not computable.

The *effective topos* Eff is a convenient home for computability theory. A statement is true in Eff if and only if it has a *computable witness*. For instance, a computable witness of a statement of the form "$\forall x.\ \exists y.\ \varphi(x, y)$" is a Turing machine which, when given an input x, computes an output y together with a computable witness for $\varphi(x, y)$.

Section 4.2.1 presents several examples to convey an intuitive understanding of truth in the effective topos; the precise translation rules are displayed in Table 4.1. A precise definition of the effective topos requires notions of category theory which we do not want to suppose here; it is included only for reference.

Introductory literature on the effective topos includes the references Hyland (1982), van Oosten (2008), Phoa (1992), and Bauer (2005).

Definition 4

1. An *assembly* is a set X together with a relation $(\Vdash_X) \subseteq \mathbb{N} \times X$ such that for every element $x \in X$, there is a number n such that $n \Vdash_X x$.
2. A *morphism of assemblies* $(X, \Vdash_X) \to (Y, \Vdash_Y)$ is a map $f : X \to Y$ which is *tracked* by a Turing machine, that is for which there exists a Turing machine M such that for any element $x \in X$ and any number n such that $n \Vdash x$, the computation $M(n)$ terminates and $M(n) \Vdash f(x)$.

A number n such that $n \Vdash_X x$ is called a *realizer* for x and can be pictured as a concrete representation of the abstract element x. The *assembly of natural numbers* is the assembly $(\mathbb{N}, =_{\mathbb{N}})$ and the *assembly of functions* $\mathbb{N} \to \mathbb{N}$ is the assembly $(X, \Vdash)$ where X is the set of computable functions $\mathbb{N} \to \mathbb{N}$ and $n \Vdash f$ if and only if the n-th Turing machine computes f. The category of assemblies is a regular category, but it is missing effective quotients. The effective topos is obtained by a suitable completion procedure:

Definition 5 The *effective topos* Eff is the ex/reg completion (as in Menni 2000, Section 3.4) of the category of assemblies.

4.2.1 Exploring the Effective Topos

Due to its computational nature, truth in the effective topos is quite different from truth in the standard topos. This section explores the following examples:

Statement	in Set	in Eff
Any natural number is prime or not prime	✓ (trivially)	✓
There are infinitely many primes	✓	✓
Any function $\mathbb{N} \to \mathbb{N}$ is constantly zero or not	✓ (trivially)	✗
Any function $\mathbb{N} \to \mathbb{N}$ is computable	✗	✓ (trivially)
Any function $\mathbb{R} \to \mathbb{R}$ is continuous	✗	✓
Markov's principle holds	✓ (trivially)	✓
Heyting arithmetic is categorical	✗	✓

Example 2 (Any natural number is prime or not.) Even without knowing what a prime number is, one can safely judge this statement to be true in the standard topos, since it is just an instance of the law of excluded middle.

By the Kripke–Joyal semantics, stating that this statement is true in the effective topos amounts to stating that there is a Turing machine which, when given a natural number n as input, terminates with a correct judgment whether n is prime or not. Such a Turing machine indeed exists – writing such a program is often a first exercise in programming courses. Hence the statement is also true in the effective topos, but for the nontrivial reason that primality can be algorithmically tested.

Example 3 (There are infinitely many primes.) A first-order formalization of this statement is "for any natural number n, there is a prime number p which is greater than n", and is known to be true in the standard topos by any of the many proofs of this fact.

Its external meaning when interpreted in the effective topos is that there exists a Turing machine M which, when given a natural number n as input, terminates with a prime number $p > n$ as output. Such a Turing machine exists, hence the statement is true in the effective topos.[9]

Example 4 (Any function $\mathbb{N} \to \mathbb{N}$ is constantly zero or not.) Precisely, the statement is

$$\forall f : \mathbb{N}^{\mathbb{N}}.\ \big((\forall n : \mathbb{N}.\ f(n) = 0) \vee \neg(\forall n : \mathbb{N}.\ f(n) = 0)\big).$$

By the law of excluded middle, this statement is trivially true in the standard topos.

[9] More precisely, the machine M should also output the description of a Turing machine which witnesses that p is prime and that $p > n$. However, the statement "p is prime and $p > n$" is $\neg\neg$-stable (even decidable), and for those statements witnesses are redundant.

Its meaning when interpreted in the effective topos is that there exists a Turing machine M which, when given the description of a Turing machine F which computes a function $f : \mathbb{N} \to \mathbb{N}$ as input, terminates with a correct judgment of whether f is the zero function or not. Such a machine M does not exist, hence the statement is false in the effective topos.

Intuitively, the issue is the following. Turing machines are able to simulate other Turing machines, hence M could simulate F on various inputs to search the list of function values $f(0), f(1), \ldots$ for a nonzero number. In case that after a certain number of steps a nonzero function value is found, the machine M can correctly output the judgment that f is not the zero function. But if the search only turned up zero values, it cannot come to any verdict – it cannot rule out that a nonzero function value will show up in the as yet unexplored part of the function.

A rigorous proof that such a machine M does not exist reduces its assumed existence to the undecidability of the halting problem.

Remark 3 Quite surprisingly, there are infinite sets X for which any flavor of constructive mathematics, in particular the kind which is valid in any topos, verifies the *omniscience principle*

$$\forall f : \mathbb{B}^X.\ \big((\exists x : X.\ f(x) = 0) \vee (\forall x : X.\ f(x) = 1)\big),$$

where $\mathbb{B} = \{0, 1\}$ is the set of booleans. This is not the case for $X = \mathbb{N}$, but it is for instance the case for the one-point compactification $X = \mathbb{N}_\infty$ of the naturals. This phenomenon has been thoroughly explored by Escardó (2013).

Example 5 (Any function $\mathbb{N} \to \mathbb{N}$ is computable.) The preceding examples give the impression that what is true in the effective topos is solely a subset of what is true in the standard topos. The example of this subsection, the so-called *formal Church–Turing thesis*, shows that the relation between the two toposes is more nuanced.

As recalled above, in the standard topos there are functions $\mathbb{N} \to \mathbb{N}$ which are not computable by a Turing machine. Cardinality arguments even show that most functions $\mathbb{N} \to \mathbb{N}$ are not computable: There are $\aleph_0^{\aleph_0} = 2^{\aleph_0}$ functions $\mathbb{N} \to \mathbb{N}$, but only $\aleph_0$ Turing machines and hence only $\aleph_0$ functions which are computable by a Turing machine.

In contrast, in the effective topos, any function $\mathbb{N} \to \mathbb{N}$ is computable by a Turing machine. The external meaning of this internal statement is that there exists a Turing machine M which, when given a description of a Turing machine F computing a function $f : \mathbb{N} \to \mathbb{N}$, outputs a description of a Turing machine computing f. It is trivial to program such a machine M: the machine M simply has to echo its input back to the user.

To avert a paradox, we should point out where the usual proof of the existence of noncomputable functions theory employs nonconstructive reasoning, for if the proof would only use intuitionistic reasoning, it would also hold internally to the effective topos, in contradiction to the fact that in the effective topos all functions $\mathbb{N} \to \mathbb{N}$ are computable.

The usual proof sets up the halting function $h : \mathbb{N} \to \mathbb{N}$, defined using the case distinction

$$h : n \mapsto \begin{cases} 1, & \text{if the } n\text{-th Turing machine terminates,} \\ 0, & \text{if the } n\text{-th Turing machine does not terminate,} \end{cases}$$

and proceeds to show that h is not computable. However, in the effective topos, this definition does not give rise to a total function from $\mathbb{N}$ to $\mathbb{N}$. The actual domain is the subset M of those natural numbers n for which the n-th Turing machine terminates or does not terminate. This condition is trivial only assuming the law of excluded middle; intuitionistically, this condition is nontrivial and cuts out a nontrivial subset of $\mathbb{N}$.

Subobjects in the effective topos are more than mere subsets; to give an element of M in the effective topos, we need not only give a natural number n such that the n-th Turing machine terminates or does not terminate, but also supply a computational witness of either case. For any particular numeral $n_0 \in \mathbb{N}$, there is such a witness (appealing to the law of excluded middle on the meta level), and hence the statement "$n_0 \in M$" holds in the effective topos. However, there is no program which could compute such witnesses for any number n, hence the statement "$\forall n : \mathbb{N}.\ n \in M$" is not true in Eff and hence the effective topos does not believe M and $\mathbb{N}$ to be the same.

Example 6 (Any function $\mathbb{R} \to \mathbb{R}$ is continuous.) In the standard topos, this statement is plainly false, with the sign and Heaviside step functions being prominent counterexamples. In the effective topos, this statement is true and independently due to Kreisel et al. (1959) and Ceĭtin (1962). A rigorous proof is not entirely straightforward (a textbook reference is Longley and Normann 2015, Theorem 9.2.1), but an intuitive explanation is as follows.

What the effective topos believes to be a real number is, from the external point of view, a Turing machine X which outputs, when called with a natural number n as input, a rational approximation $X(n)$. These approximations are required to be *consistent* in the sense that $|X(n) - X(m)| \leq 2^{-n} + 2^{-m}$. Intuitively, such a machine X denotes the real number $\lim_{n\to\infty} X(n)$, and each approximation $X(n)$ must be within 2^{-n} of the limit.

A function $f : \mathbb{R} \to \mathbb{R}$ in the effective topos is therefore given by a Turing machine M which, when given the description of such a Turing machine X as input, outputs the description of a similar such Turing machine Y. To compute a rational approximation $Y(n)$, the machine Y may simulate X and can therefore determine arbitrarily many rational approximations $X(m)$. However, within a finite amount of time, the machine Y can only learn finitely many such approximations. Hence a function such as the sign function, for which even rough rational approximations of $\operatorname{sgn}(x)$ require infinite precision in the input x, does not exist in the effective topos.

Example 7 (Markov's principle holds.) Markov's principle is the following statement:

$$\forall f : \mathbb{N}^{\mathbb{N}}.\ \big((\neg\neg\exists n : \mathbb{N}.\ f(n) = 0) \Longrightarrow \exists n : \mathbb{N}.\ f(n) = 0\big). \tag{MP}$$

It is an instance of the law of double negation elimination and hence trivially true in the standard topos, at least if we subscribe to classical logic on the meta level. A useful consequence of Markov's principle is that Turing machines which do not run forever (that is, which do *not not* terminate) actually terminate; this follows by applying Markov's principle to the function $f : \mathbb{N} \to \mathbb{N}$ where $f(n)$ is zero or one depending on whether a given Turing machine has terminated within n computational steps or not.

The effective topos inherits Markov's principle from the meta level: The statement "Eff $\models$ (MP)" means that there is a Turing machine M which, when given the description of a Turing machine F computing a function $f : \mathbb{N} \to \mathbb{N}$ as input, outputs the description of a Turing machine S_F which, when given a witness of "$\neg\neg\exists n : \mathbb{N}.\ f(n) = 0$", outputs a witness of "$\exists n : \mathbb{N}.\ f(n) = 0$" (up to trivial conversions, this is a number n such that $f(n) = 0$).

By the translation rules listed in Table 4.1, a number e realizes "$\neg\neg\exists n : \mathbb{N}.\ f(n) = 0$" if and only if it is *not not* the case that there is some number e' such

Table 4.1 A (fragment of) the translation rules defining the meaning of statements internal to the effective topos

Eff $\models \varphi$	iff there is a natural number e such that $e \Vdash \varphi$.

A number e such that $e \Vdash \varphi$ is called a *realizer* for φ. It is the precise version of what is called *computational witness* in the main text. In the following, we write "$e \cdot n \downarrow$" to mean that the e-th Turing machine terminates on input n, and in this case denote the result by "$e \cdot n$". No separate clause for negation is listed, as "$\neg\varphi$" is an abbreviation for "$(\varphi \Rightarrow \bot)$".

$e \Vdash s = t$	iff $s = t$.
$e \Vdash \top$	iff $1 = 1$.
$e \Vdash \bot$	iff $1 = 0$.
$e \Vdash (\varphi \wedge \psi)$	iff $e \cdot 0 \downarrow$ and $e \cdot 1 \downarrow$ and $e \cdot 0 \Vdash \varphi$ and $e \cdot 1 \Vdash \psi$.
$e \Vdash (\varphi \vee \psi)$	iff $e \cdot 0 \downarrow$ and $e \cdot 1 \downarrow$ and if $e \cdot 0 = 0$ then $e \cdot 1 \Vdash \varphi$, and if $e \cdot 0 \neq 0$ then $e \cdot 1 \Vdash \psi$.
$e \Vdash (\varphi \Rightarrow \psi)$	iff for any number $r \in \mathbb{N}$ such that $r \Vdash \varphi$, $e \cdot r \downarrow$ and $e \cdot r \Vdash \psi$.
$e \Vdash (\forall n : \mathbb{N}.\ \varphi(n))$	iff for any number $n_0 \in \mathbb{N}$, $e \cdot n_0 \downarrow$ and $e \cdot n_0 \Vdash \varphi(n_0)$.
$e \Vdash (\exists n : \mathbb{N}.\ \varphi(n))$	iff $e \cdot 0 \downarrow$ and $e \cdot 1 \downarrow$ and $e \cdot 1 \Vdash \varphi(e \cdot 0)$.
$e \Vdash (\forall f : \mathbb{N}^{\mathbb{N}}.\ \varphi(f))$	iff for any function $f_0 : \mathbb{N} \to \mathbb{N}$ and any number r_0 such that f_0 is computed by the r_0-th Turing machine, $e \cdot r_0 \downarrow$ and $e \cdot r_0 \Vdash \varphi(f_0)$.
$e \Vdash (\exists f : \mathbb{N}^{\mathbb{N}}.\ \varphi(f))$	iff $e \cdot 0 \downarrow$ and $e \cdot 1 \downarrow$ and the $(e \cdot 0)$-th Turing machine computes a function $f_0 : \mathbb{N} \to \mathbb{N}$ and $e \cdot 1 \Vdash \varphi(f_0)$.

that e' realizes "$\exists n : \mathbb{N}.\ f(n) = 0$". Hence, if "$\exists n : \mathbb{N}.\ f(n) = 0$" is realized at all, then any number is a witness of "$\neg\neg\exists n : \mathbb{N}.\ f(n) = 0$".

As a consequence, the input given to the machine S_F is entirely uninformative and S_F cannot make direct computational use of it. But its existence ensures that an *unbounded search* will not fail (and hence succeed, by an appeal to Markov's principle on the meta level): The machine S_F can simulate F to compute the values $f(0)$, $f(1)$, $f(2)$, ... in turn, and stop with output n as soon as it determines that some function value $f(n)$ is zero.

Example 8 (Heyting arithmetic is categorical.) In addition to the standard model $\mathbb{N}$, the standard topos contains uncountably many nonstandard models of Peano arithmetic (at least if we assume a classical meta level). By a theorem of van den Berg and van Oosten (2011), the situation is quite different in the effective topos:

1. Heyting arithmetic, the intuitionistic cousin of Peano arithmetic, is categorical in the sense that it has exactly one model up to isomorphism, namely $\mathbb{N}$.
2. In fact, even the finitely axiomatizable subsystem of Heyting arithmetic where the induction scheme is restricted to Σ_1-formulas has exactly one model up to isomorphism, again $\mathbb{N}$. As a consequence, Heyting arithmetic is finitely axiomatizable.
3. Peano arithmetic is "quasi-inconsistent" in that it does not have any models, for any model of Peano arithmetic would also be a model of Heyting arithmetic, but the only model of Heyting arithmetic is $\mathbb{N}$ and $\mathbb{N}$ does not validate the theorem "any Turing machine terminates or does not terminate" of Peano arithmetic.

As a consequence, Gödel's completeness theorem fails in the effective topos: In the effective topos, Peano arithmetic is consistent (because it is equiconsistent to Heyting arithmetic, which has a model) but does not have a model.

Statement (1) is reminiscent of the fact due to Tennenbaum (1959) that no nonstandard model of Peano arithmetic in the standard topos is computable.

4.2.2 Variants of the Effective Topos

The effective topos belongs to a wider class of *realizability toposes*. These can be obtained by repeating the construction of the effective topos with any other reasonable model of computation in place of Turing machines. The resulting toposes will in general not be equivalent and reflect higher-order properties of the employed models. Two of these further toposes are of special philosophical interest.

Hypercomputation Firstly, in place of ordinary Turing machines, one can employ the *infinite-time Turing machines* pioneered by Hamkins and Lewis (2000). These machines model *hypercomputation* in that they can run for "longer than infinity"; more precisely, their computational steps are indexed by the ordinal numbers instead of the natural numbers. For instance, an infinite-time Turing machine can trivially decide the twin prime conjecture, by simply walking along the natural number line

and recording any twin primes it finds. Then, on day ω, it can observe whether it has found infinitely many twins or not.

In the realizability topos constructed using infinite-time Turing machines, the full law of excluded middle still fails, but some instances which are wrong in the effective topos do hold in this topos. For example, the instance "any function $\mathbb{N} \to \mathbb{N}$ is the zero function or not" does: Its external meaning is that there is an infinite-time Turing machine M which, when given the description of an infinite-time Turing machine F computing a function $f : \mathbb{N} \to \mathbb{N}$ as input, terminates (at some ordinal time step) with a correct judgment of whether f is the zero function or not. Such a machine M indeed exists: It simply has to simulate F on all inputs $0, 1, \ldots$ and check whether one of the resulting function values is not zero. This search will require a transfinite amount of time (not least because simulating F on just one input might require a transfinite amount of time), but infinite-time Turing machines are capable of carrying out this procedure.

The realizability topos given by infinite-time Turing machines provides an intriguing environment challenging many mathematical intuitions shaped by classical logic. For instance, while from the point of view of this topos the reals are still uncountable in the sense that there is no surjection $\mathbb{N} \to \mathbb{R}$, there is an injection $\mathbb{R} \to \mathbb{N}$ (Bauer, 2015, Section 4).[10]

Machines of the physical world A second variant of the effective topos is obtained by using machines of the physical world instead of abstract Turing machines. In doing so, we of course leave the realm of mathematics, as real-world machines are not objects of mathematical study. But it is still interesting to see which

[10] What the realizability topos given by infinite-time Turing machines believes to be a real number is, from the external point of view, an infinite-time Turing machine X which outputs, when called with a natural number n as input, a rational approximation $X(n)$. As with the original effective topos, these approximations have to be consistent in the sense that $|X(n) - X(m)| \leq 2^{-n} + 2^{-m}$, and two such machines X, X' represent the same real iff $|X(n) - X'(m)| \leq 2^{-n} + 2^{-m}$ for all natural numbers n, m.

A map $\mathbb{R} \to \mathbb{N}$ in this topos is hence given by an infinite-time Turing machine M which, when given the description of such an infinite-time Turing machine X as input, outputs a certain natural number $M(X)$. If X and X' represent the same real, then $M(X)$ has to coincide with $M(X')$. This map $\mathbb{R} \to \mathbb{N}$ is injective iff conversely $M(X) = M(X')$ implies that X and X' represent the same real.

We can program such a machine M as follows: Read the description of an infinite-time Turing machine X representing a real number as input. Then simulate, in a dovetailing fashion, all infinite-time Turing machines and compare their outputs with the outputs of X. As soon as a machine X' is found which happens to terminate on all inputs in such a way that $|X(n) - X'(m)| \leq 2^{-n} + 2^{-m}$ for all natural numbers n, m, output the number of this machine (in the chosen enumeration of all infinite-time Turing machines) and halt.

The number $M(X)$ computed by M depends on the input/output behaviour of X, the chosen ordering of infinite-time Turing machines, and on details of the interleaving simulation and the comparison procedure – but it does not depend on the implementation of X or on its specific choice of rational approximations $X(n)$. The search terminates since there is at least one infinite-time Turing machine which represents the same real number as X does, namely X itself.

commitments about the nature of the physical world imply which internal statements of the resulting topos.

For instance, Bauer (2012) showed that inside this topos any function $\mathbb{R} \to \mathbb{R}$ is continuous if, in the physical world, only finitely many computational steps can be carried out in finite time and if it is possible to form tamper-free private communication channels.

4.3 Toposes of Sheaves, a Convenient Home for Local Truth

Associated to any topological space X (such as Euclidean space), there is the *topos of sheaves over* X, $\mathrm{Sh}(X)$. To a first approximation, a statement is true in $\mathrm{Sh}(X)$ if and only if it "holds locally on X"; what $\mathrm{Sh}(X)$ believes to be a set is a "continuous family of sets, one set for each point of X". The precise rules of the Kripke–Joyal semantics of $\mathrm{Sh}(X)$ are listed in Table 4.2.

Just as the effective topos provides a coherent setting for studying computability using a naive element-based language, the sheaf topos $\mathrm{Sh}(X)$ provides a coherent setting for studying continuous X-indexed families of objects (sets, numbers, functions) as if they were single objects.

Sheaf toposes take up a special place in the history of topos theory: If the base X is allowed to be a *site* instead of a topological space, the resulting toposes constitute the large class of Grothendieck toposes, the original notion of toposes. Categorically, the passage from topological spaces to sites is rather small, but the

Table 4.2 A (fragment of) the translation rules defining the meaning of statements internal to $\mathrm{Sh}(X)$, the topos of sheaves over a topological space X

$\mathrm{Sh}(X) \models \varphi$	iff $X \models \varphi$.
$U \models a = b$	iff $a = b$ on U.
$U \models \top$	is true for any open U.
$U \models \bot$	iff U is the empty open.
$U \models (\varphi \wedge \psi)$	iff $U \models \varphi$ and $U \models \psi$.
$U \models (\varphi \vee \psi)$	iff there is an open covering $U = \bigcup_i U_i$ such that, for each index i, $U_i \models \varphi$ or $U_i \models \psi$.
$U \models (\varphi \Rightarrow \psi)$	iff for any open $V \subseteq U$, $V \models \varphi$ implies $V \models \psi$.
$U \models (\forall a : \mathbb{R}.\ \varphi(a))$	iff for any open $V \subseteq U$ and any continuous function $a_0 : V \to \mathbb{R}$, $V \models \varphi(a_0)$.
$U \models (\exists a : \mathbb{R}.\ \varphi(a))$	iff there is an open covering $U = \bigcup_i U_i$ such that, for each index i, there exists a continuous function $a_0 : U_i \to \mathbb{R}$ with $U_i \models \varphi(a_0)$.

resulting increase in flexibility is substantial and fundamental to modern algebraic geometry.

4.3.1 A Geometric Interpretation of Double Negation

In intuitionistic logic, the double negation $\neg\neg\varphi$ of a statement φ is a slight weakening of φ; while $(\varphi \Rightarrow \neg\neg\varphi)$ is an intuitionistic tautology, the converse can only be shown for some specific statements. The internal language of $\mathrm{Sh}(X)$ gives geometric meaning to this logical peculiarity.

Namely, it is an instructive exercise that $\mathrm{Sh}(X) \models \neg\neg\varphi$ is equivalent to the existence of a *dense open* U of X such that $U \models \varphi$. If $\mathrm{Sh}(X) \models \varphi$, that is if $X \models \varphi$, then there obviously exists such a dense open, namely X itself; however the converse usually fails.

The only case that the law of excluded middle does hold internally to $\mathrm{Sh}(X)$ is when the only dense open of X is X itself; assuming classical logic in the metatheory, this holds if and only if every open is also closed. This is essentially only satisfied if X is discrete.

An important special case is when X is the one-point space. In this case $\mathrm{Sh}(X)$ is equivalent (as categories and hence toposes) to the standard topos. To the extent that mathematics within $\mathrm{Sh}(X)$ can be described as "mathematics over X", this observation justifies the slogan that "ordinary mathematics is mathematics over the point".

4.3.2 Reifying Continuous Families of Real Numbers as Single Real Numbers

As detailed in Example 6, what the effective topos believes to be a real number is actually a Turing machine computing arbitrarily-good consistent rational approximations. A similarly drastic shift in meaning, though in an orthogonal direction, occurs with $\mathrm{Sh}(X)$. What $\mathrm{Sh}(X)$ believes to be a (Dedekind) real number a is actually a continuous family of real numbers on X, that is, a continuous function $a : X \to \mathbb{R}$ (Johnstone, 2002, Corollary D4.7.5).

Such a function is everywhere positive on X if and only if, from the internal point of view of $\mathrm{Sh}(X)$, the number a is positive; it is everywhere zero if and only if, internally, the number a is zero; and it is everywhere negative if and only if, internally, the number a is negative.

The law of trichotomy, stating that any real number is either negative, zero or positive, generally fails in $\mathrm{Sh}(X)$. By the Kripke–Joyal semantics, the external meaning of the internal statement "$\forall a : \mathbb{R}.\ a < 0 \vee a = 0 \vee a > 0$" is that for any continuous function $a : U \to \mathbb{R}$ defined on any open U of X, there is an open

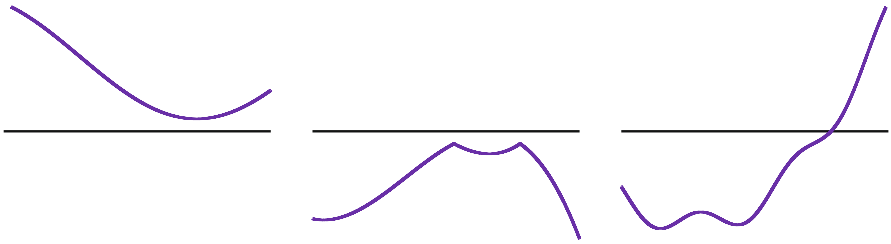


Fig. 4.2 Three examples of what the topos $\mathrm{Sh}(X)$ believes to be a single real number, where the base space X is the unit interval. (**a**) A positive real number. (**b**) A negative real number. (**c**) A number which is neither negative nor zero nor positive. Externally speaking, there is no covering of the unit interval by opens on which the depicted function a is either everywhere negative, everywhere zero or everywhere positive

covering $U = \bigcup_i U_i$ such that on each member U_i of this covering, the function a is either everywhere negative on U_i, everywhere zero on U_i or everywhere positive on U_i. But this statement is, for most base spaces X, false. Figure 4.2c shows a counterexample.

The weaker statement "for any real number a it is *not not* the case that $a < 0$ or $a = 0$ or $a > 0$" does hold in $\mathrm{Sh}(X)$, for this statement is a theorem of intuitionistic calculus. Its meaning is that there exists a dense open U such that U can be covered by opens on which a is either everywhere negative, everywhere zero or everywhere positive. In the example given in Fig. 4.2c, this open U could be taken as X with the unique zero of a removed.

4.3.3 *Reifying Continuous Families of Real Functions as Single Real Functions*

Let $(f_x)_{x \in X}$ be a continuous family of continuous real-valued functions; that is, not only should each of the individual functions $f_x : \mathbb{R} \to \mathbb{R}$ be continuous, but the joint map $\mathbb{R} \times X \to \mathbb{R}$, $(a, x) \mapsto f_x(a)$ should also be continuous. (This stronger condition implies continuity of the individual functions.) From the point of view of $\mathrm{Sh}(X)$, this family looks like a single continuous function $f : \mathbb{R} \to \mathbb{R}$.

The internal statement "$f(-1) < 0$" means that $f_x(-1) < 0$ for all $x \in X$, and similarly so for being positive. More generally, if a and b are continuous functions $X \to \mathbb{R}$ (hence real numbers from the internal point of view), the internal statement "$f(a) < b$" means that $f_x(a(x)) < b(x)$ for all $x \in X$.

The internal statement "f possesses a zero", that is "there exists a number a such that $f(a) = 0$", means that all the functions f_x each possess a zero and that moreover, these zeros can locally be picked in a continuous fashion. More precisely, this statement means that there is an open covering $X = \bigcup_i U_i$ such that, for each

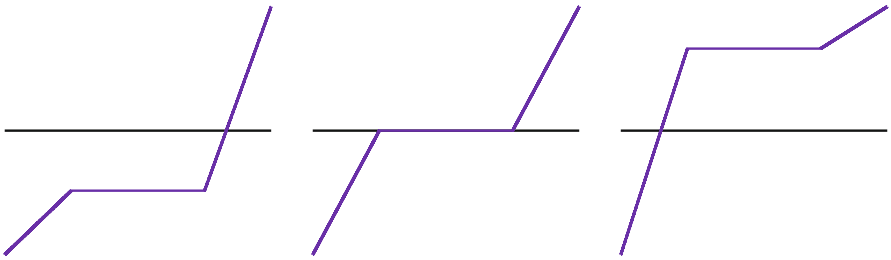


Fig. 4.3 Three members f_{x_0}, f_{x_1}, f_{x_2} of a continuous family $(f_x)_{x\in X}$ of continuous functions $f_x : \mathbb{R} \to \mathbb{R}$. The parameter space is $X = [0, 1]$ (not shown). The functions f_x are obtained by moving the horizontal plateau up or down. The leftmost depicted member f_{x_0} has a unique zero, and there is an open neighborhood U of x_0 on which zeros of the functions f_x, $x \in U$ can be picked continuously. The same is true for x_2 (right figure). However, there is no such neighborhood of that particular parameter value x_1 for which the horizontal plateau lies on the x-axis (middle figure)

index i, there is a continuous function $a : U_i \to \mathbb{R}$ such that $f_x(a(x)) = 0$ for all $x \in U_i$. (On overlaps $U_i \cap U_j$, the zero-picking functions a need not agree.)

Example 9 From these observations we can deduce that the intermediate value theorem of undergraduate calculus does in general not hold in $\mathrm{Sh}(X)$ and hence does not allow for an intuitionistic proof. The intermediate value theorem states: "If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function such that $f(-1) < 0$ and $f(1) > 0$, there exists a number a such that $f(a) = 0$." The external meaning of this statement is that for any continuous family $(f_x)_x$ of continuous functions with $f_x(-1) < 0$ and $f_x(1) > 0$ for all $x \in X$, it is locally possible to pick zeros of the family in a continuous fashion. Figure 4.3 shows a counterexample to this claim.

In contrast, the intermediate value theorem for (strictly) monotone functions does have an intuitionistic proof and hence applies in the internal universe of $\mathrm{Sh}(X)$. Thus for any continuous family $(f_x)_x$ of continuous monotone functions with $f_x(-1) < 0$ and $f_x(1) > 0$ for all $x \in X$, it is locally possible to pick zeros of the family in a continuous fashion.[11]

Example 10 The fundamental theorem of algebra generally fails in $\mathrm{Sh}(X)$, even for quadratic polynomials. What $\mathrm{Sh}(X)$ believes to be a (Dedekind) complex number is externally a continuous function $X \to \mathbb{C}$. Let X be the complex plane. Then the identity function id_X is a single complex number from the internal point of view of $\mathrm{Sh}(X)$. The fundamental theorem of algebra would predict "$\mathrm{Sh}(X) \models \exists a : \mathbb{C}.\ a^2 - \mathrm{id}_X = 0$", hence that there is an open covering $X = \bigcup_i U_i$ such that on each open U_i, there is a continuous function $a : U_i \to \mathbb{C}$ such that $a(z)^2 - z = 0$

[11] While the Kripke–Joyal translation of "$\exists$" is by definition *local existence*, one can show that the Kripke–Joyal translation of "$\exists!$" is *unique existence on all opens*, in particular *unique global existence*. Because the conclusion of the intermediate value theorem for monotone functions can be strengthened from "has a zero" to "has a unique zero", this observation shows that the zeros can even globally be picked in a continuous fashion.

for all $z \in U_i$. However, it is a basic fact of complex analysis that such a function does not exist if $0 \in U_i$.

Example 11 The standard proof of Banach's fixed point theorem employs only intuitionistic reasoning, hence applies internally to $\mathrm{Sh}(X)$. Interpreting the internal Banach fixed point theorem by the Kripke–Joyal translation rules yields the statement that fixed points of continuous families of contractions depend continuously on parameters.

4.4 Toposes Adapted to Synthetic Differential Geometry

The idea of *infinitesimal numbers* – numbers which can be pictured as lying between $-\frac{1}{n}$ and $\frac{1}{n}$ for any natural number n (though this intuition will not serve as their formal definition in this text) – has a long and rich history. They are not part of today's standard setup of the reals, but they are still intriguing as a calculational tool and as a device to bring mathematical intuition and mathematical formalism closer together.

For instance, employing numbers ε such that $\varepsilon^2 = 0$, we can compute derivatives blithely as follows, without requiring the notion of limits:

$$\begin{aligned}(x+\varepsilon)^2 - x^2 &= x^2 + 2x\varepsilon + \varepsilon^2 - x^2 = 2x\varepsilon \\ (x+\varepsilon)^3 - x^3 &= x^3 + 3x^2\varepsilon + 3x\varepsilon^2 + \varepsilon^3 - x^3 = 3x^2\varepsilon\end{aligned} \qquad (\star)$$

In each case, the derivative is visible as the coefficient of ε in the result. A further example is from geometry: Having a nontrivial set Δ of infinitesimal numbers available allows us to define a *tangent vector* to a manifold M to be a map $\gamma : \Delta \to M$. This definition precisely captures the intuition that a tangent vector is an infinitesimal curve.

4.4.1 Hyperreal Numbers

There are several ways of introducing infinitesimals into rigorous mathematics. One is Robinson's *nonstandard analysis*, where we enlarge the field $\mathbb{R}$ of real numbers to a field ${}^\star\mathbb{R}$ of *hyperreal numbers* by means of a non-principal ultrafilter.

The hyperreals contain an isomorphic copy of the ordinary reals as the so-called *standard elements*, and they also contain infinitesimal numbers and their inverses, transfinite numbers. Additionally, they support a powerful *transfer principle*: Any statement which does not refer to standardness is true for the hyperreals if and only if it is true for the ordinary reals.

In the "if" direction, the transfer principle is useful for importing knowledge about the ordinary reals into the hyperreal realm. For instance, addition of hyperreals is commutative because addition of reals is. By the "only if" direction, a theorem established for the hyperreals also holds for the ordinary reals. In this way, the infinitesimal numbers of nonstandard analysis can be viewed as a convenient fiction, generating a conservative extension of the usual setup of mathematics.

There is a growing body of research in mathematics which employs hyperreal numbers in this sense. To exemplarily cite just one example, a recent application of nonstandard analysis in symplectic geometry is due to Fabert (2015a,b), who verified an infinite-dimensional analogue of the Arnold conjecture.

However, the realization of the fiction of infinitesimal numbers in nonstandard analysis crucially rests on a non-principal ultrafilter, whose existence requires principles which go beyond the means of Zermelo–Fraenkel set theory ZF.[12] Non-principal ultrafilters cannot be described in explicit terms, and they are also not at all canonical structures: ZFC proves that there are $2^{2^{\aleph_0}}$ many, see Pospíšil (1937).

A practical consequence of this nonconstructivity is that it can be hard to unwind proofs which employ hyperreal numbers to direct proofs, and even where possible there is no general procedure for doing so. For instance, Fabert has not obtained a direct proof of his result, and not for the lack of trying (personal communication).

4.4.2 Topos-Theoretic Alternatives to the Hyperreal Numbers

Topos theory provides several constructive alternatives for realizing infinitesimals. One such is "cheap nonstandard analysis" by Tao (2012). It is to Robinson's nonstandard analysis what potential infinity is to actual infinity: Instead of appealing to the axiom of choice to obtain a completed ultrafilter, cheap nonstandard analysis constructs larger and larger approximations to an ideal ultrafilter on the go.

The following section presents a (variant of a) topos used in *synthetic differential geometry* as discussed by Kock (2006, 2020). This subject is a further topos-theoretic approach to infinitesimals which is suited to illustrate the philosophy of toposes as lenses. A major motivation for the development of synthetic differential

[12] A hyperreal number is represented by an infinite sequence $(x_0, x_1, x_2, \ldots)$ of ordinary real numbers. For instance, the sequence $(1, 1, 1, \ldots)$ represents the hyperreal version of the number 1, the sequence $(1, \frac{1}{2}, \frac{1}{3}, \ldots)$ represents an infinitesimal number and its inverse $(1, 2, 3, \ldots)$ represents a transfinite number. The sequence $(1, 1, 1, \ldots)$ is deemed positive, and so is $(-1, 1, 1, 1, \ldots)$, which differs from the former only in finitely many places. But should $(1, -1, 1, -1, \ldots)$ be deemed positive or negative? Whatever the answer, our decision has consequences for other sequences. For instance $(-1, 1, -1, 1, \ldots)$ should be assigned the opposite sign and $(\tan(1), \tan(-1), \tan(1), \tan(-1), \ldots)$ the same. A non-principal ultrafilter is a set-theoretic gadget which fixes all such decisions once and for all in a coherent manner. Having such an ultrafilter available, a sequence $(x_0, x_1, x_2, \ldots)$ is deemed positive if and only if the set $\{i \in \mathbb{N} \mid x_i > 0\}$ is part of the ultrafilter.

geometry was to devise a rigorous context in which the writings of Sophus Lie, who freely employed infinitesimals in his seminal works, can be effortlessly interpreted, staying close to the original and requiring no coding.

4.4.3 The Zariski Topos

The starting point is the observation that while the field $\mathbb{R}$ of ordinary real numbers does not contain infinitesimals (except for zero), the ring $\mathbb{R}[\varepsilon]/(\varepsilon^2)$ of *dual numbers* does. This ring has the cartesian product $\mathbb{R} \times \mathbb{R}$ as its underlying set and the ring operations are defined such that $\varepsilon^2 = 0$, where $\varepsilon := \langle 0, 1\rangle$:

$$\langle a, b\rangle + \langle a', b'\rangle := \langle a + a', b + b'\rangle \qquad \langle a, b\rangle \cdot \langle a', b'\rangle := \langle aa', ab' + a'b\rangle$$

We write $\langle a, b\rangle$ more clearly as $a + b\varepsilon$.

The flavor of infinitesimal numbers supported by $\mathbb{R}[\varepsilon]/(\varepsilon^2)$ are the *nilsquare numbers*, numbers which square to zero. The numbers $b\varepsilon$ with $b \in \mathbb{R}$ are nilsquare in $\mathbb{R}[\varepsilon]/(\varepsilon^2)$, and they are sufficient to rigorously reproduce derivative computations of polynomials such as (⋆).

However, the dual numbers are severely lacking in other aspects. Firstly, they do not contain any nilcube numbers which are not already nilsquare. These are required in order to extend calculations like (⋆) to second derivatives, as in

$$(x + \varepsilon)^3 - x^3 = 3x^2\varepsilon + \frac{1}{2!}6x\varepsilon^2.$$

Secondly, the dual numbers contain, up to scaling, only a single infinitesimal number. Further independent infinitesimals are required in order to deal with functions of several variables, as in

$$f(x + \varepsilon, y + \varepsilon') - f(x, y) = D_x f(x, y)\varepsilon + D_y f(x, y)\varepsilon'.$$

Thirdly, and perhaps most importantly, the ring of dual numbers fails to be a field. The only invertible dual numbers are the numbers of the form $a+b\varepsilon$ with a invertible in the reals; it is not true that any nonzero dual number is invertible.

The first deficiency could be fixed by passing from $\mathbb{R}[\varepsilon]/(\varepsilon^2)$ to $\mathbb{R}[\varepsilon]/(\varepsilon^3)$ (a ring whose elements are triples and whose ring operations are defined such that $\langle 0, 1, 0\rangle^3 = 0$) and the second by passing from $\mathbb{R}[\varepsilon]/(\varepsilon^2)$ to $\mathbb{R}[\varepsilon, \varepsilon']/(\varepsilon^2, \varepsilon'^2, \varepsilon\varepsilon')$. In a sense, both of these proposed replacements are better *stages* than the basic ring $\mathbb{R}[\varepsilon]/(\varepsilon^2)$ or even $\mathbb{R}$ itself. However, similar criticisms can be mounted against any of these better stages, and the problem that all these substitutes are not fields persists.

Introducing the topos The *Zariski topos of* $\mathbb{R}$, $\mathrm{Zar}(\mathbb{R})$, meets all of these challenges. It contains a ring $\mathbb{R}^\sim$, the so-called *ring of smooth numbers*, which reifies

Table 4.3 A (fragment of) the Kripke–Joyal translation rules of the Zariski topos $\mathrm{Zar}(\mathbb{R})$

$\mathrm{Zar}(\mathbb{R}) \models \varphi$	iff $\mathbb{R} \models \varphi$.
$A \models s = t$	iff $s = t$ as elements of A.
$A \models \top$	iff $1 = 1$ in A.
$A \models \bot$	iff $1 = 0$ in A.
$A \models (\varphi \wedge \psi)$	iff $A \models \varphi$ and $A \models \psi$.
$A \models (\varphi \vee \psi)$	iff there exists a partition $1 = f_1 + \cdots + f_n \in A$ such that, for each index i, $A[f_i^{-1}] \models \varphi$ or $A[f_i^{-1}] \models \psi$.
$A \models (\varphi \Rightarrow \psi)$	iff for any finitely presented A-algebra B, $B \models \varphi$ implies $B \models \psi$.
$A \models (\forall x : \mathbb{R}^{\sim}.\ \varphi(x))$	iff for any finitely presented A-algebra B and any element $x_0 \in B$, $B \models \varphi(x_0)$.
$A \models (\exists x : \mathbb{R}^{\sim}.\ \varphi(x))$	iff there exists a partition $1 = f_1 + \cdots + f_n \in A$ such that, for each index i, there is an element $x_0 \in A[f_i^{-1}]$ such that $A[f_i^{-1}] \models \varphi(x_0)$.

A *covering* of an $\mathbb{R}$-algebra A is a finite family of A-algebras of the form $(A[f_i^{-1}])_{i=1,\ldots,n}$ such that $1 = f_1 + \cdots + f_n \in A$.

the real numbers, the dual numbers, the two proposed better stages and indeed any finitely presented $\mathbb{R}$-algebra into a single coherent entity. The Kripke–Joyal translations rules of $\mathrm{Zar}(\mathbb{R})$ are listed in Table 4.3. Any evaluation of an internal statement starts out with the most basic stage of all, the ordinary reals $\mathbb{R}$; then, during the course of evaluation, the current stage is successively refined to better stages (further finitely presented $\mathbb{R}$-algebras).

For instance, universal quantification "$\forall x : \mathbb{R}^{\sim}$" not only refers to all elements of the current stage, but also to any elements of any refinement of the current stage. Similarly, negation "$\neg\varphi$" does not only mean that φ would imply $\bot$ in the current stage, but also that it does so at any later stage.

For reference purposes only, we include the precise definition of the Zariski topos.

Definition 6 The *Zariski topos of* $\mathbb{R}$, $\mathrm{Zar}(\mathbb{R})$, is a certain full subcategory of the category of functors from finitely presented $\mathbb{R}$-algebras to sets, namely of the Zariski sheaves. Such a functor is a *Zariski sheaf* if and only if, for any covering $(A[f_i^{-1}])_i$ of any finitely presented $\mathbb{R}$-algebra A (this notion is defined in Table 4.3), the diagram

$$F(A) \to \prod_i F(A[f_i^{-1}]) \rightrightarrows \prod_{j,k} F(A[(f_j f_k)^{-1}])$$

is an equalizer diagram. The object $\mathbb{R}^{\sim}$ of $\mathrm{Zar}(\mathbb{R})$ is the tautologous functor $A \mapsto A$.

Properties of the smooth numbers As a concrete example, the Kripke–Joyal translation of the statement that $\mathbb{R}^\sim$ is a field,

$$\mathrm{Zar}(\mathbb{R}) \models \forall x : \mathbb{R}^\sim.\ (\neg(x = 0) \Rightarrow \exists y : \mathbb{R}^\sim.\ xy = 1),$$

is this:

> For any stage A and any element $x \in A$,
> for any later stage B of A,
> if for any later stage C of B
> in which $x = 0$ holds
> also $1 = 0$ holds,
> then B can be covered by later stages C_i such that,
> for each index i, there is an element $y \in C_i$ with $xy = 1$ in C_i.

And indeed, this statement is true. Let a stage A (a finitely presented $\mathbb{R}$-algebra) and an element $x \in A$ be given. Let B be any later stage of A (any finitely presented A-algebra – such an algebra is also finitely presented as an $\mathbb{R}$-algebra). Assume that for any later stage C of B in which $x = 0$ holds also $1 = 0$ holds. Trivially, $x = 0$ holds in the particular refinement $C := B/(x)$. Hence $1 = 0$ holds in C. By elementary algebra, this means that x is invertible in B. Hence the conclusion holds for the singleton covering of B given by $C_1 := B$.

Remark 4 The Zariski topos can also be set up with an arbitrary commutative ring S in place of $\mathbb{R}$. The resulting topos $\mathrm{Zar}(S)$ contains a mirror image $S^\sim$ of S, a reification of all finitely presented S-algebras into a single entity. The computation we just carried out also applies in this more general context and shows that $S^\sim$ is a field. It is in this sense that the topos $\mathrm{Zar}(S)$ provides a lens through which S looks like a field.

A small variant of this lens has been used to give a new proof of *Grothendieck's generic freeness lemma*, a fundamental theorem in algebraic geometry about the free locus of certain sheaves. The new proof uses the lens to reduce to the case of fields, where the claim is trivial (Blechschmidt, 2017, Section 11.5), and improves in length on all previously known proofs, even if the topos machinery is eliminated by unrolling the appropriate definitions as in Blechschmidt (2018).[13]

[13] This contribution is not the proper place for an exposition of Grothendieck's generic freeness lemma, but some aspects can already be appreciated on a syntactical level. Grothendieck's generic freeness lemma states that any finitely generated sheaf of modules on a reduced scheme is finite locally free *on a dense open*. By employing the internal language, this statement is reduced to the following fact of intuitionistic linear algebra: Any finitely generated module over a field is *not not* finite free.

Within Zar($\mathbb{R}$), we can construct the set $\Delta := \{\varepsilon : \mathbb{R}^\sim \mid \varepsilon^2 = 0\}$ of nilsquare numbers. Then $\mathbb{R}^\sim$ validates the following laws:

1. Law of cancellation: $\forall x : \mathbb{R}^\sim.\ \forall y : \mathbb{R}^\sim.\ \big((\forall \varepsilon : \Delta.\ x\varepsilon = y\varepsilon) \Rightarrow x = y\big)$
2. Axiom of micro-affinity: $\forall f : (\mathbb{R}^\sim)^\Delta.\ \exists! a : \mathbb{R}^\sim.\ \forall \varepsilon : \Delta.\ f(\varepsilon) = f(0) + a\varepsilon$

The unique number a in the axiom of micro-affinity deserves to be called "$f'(0)$"; this is how we synthetically define the derivative in synthetic differential geometry. (However, despite these properties the Zariski topos is not yet *well-adapted* to synthetic differential geometry in the sense of Definition 7 below.)

Having motivated the Zariski topos by the desire to devise a universe with infinitesimals, the actual ontological status of the infinitesimal numbers in the Zariski topos is more nuanced. The law of cancellation implies that, within Zar($\mathbb{R}$), it is not the case that zero is the only nilsquare number. However, this does not mean that there actually *is* a nilsquare number in $\mathbb{R}^\sim$. In fact, any nilsquare number cannot be nonzero, as nonzero numbers are invertible while nilsquare numbers are not. Hence any nilsquare number in $\mathbb{R}^\sim$ is *not not* zero. This state of affairs is only possible in an intuitionistic context.

Remark 5 The ring $\mathbb{R}^\sim$ of smooth numbers does not coincide with the Cauchy reals, the Dedekind reals or indeed any well-known construction of the reals within Zar($\mathbb{R}$). This observation explains why $\mathbb{R}^\sim$ can satisfy the law of cancellation even though it is an intuitionistic theorem that the only nilsquare number in any flavor of the reals is zero.

4.4.4 Well-Adapted Models

The Zariski topos of $\mathbb{R}$ allows to compute with infinitesimals in a satisfying manner. However it is not suited as a home for synthetic differential geometry, a first indication being that in Zar($\mathbb{R}$), any function $\mathbb{R}^\sim \to \mathbb{R}^\sim$ is a polynomial function. Hence important functions such as the exponential function do not exist in Zar($\mathbb{R}$). More comprehensively, the Zariski topos is not a well-adapted model in the sense of the following definition.

Definition 7 A *well-adapted model* of synthetic differential geometry is a topos $\mathcal{E}$ together with a ring $\mathbb{R}^\sim$ in $\mathcal{E}$ such that:

1. The ring $\mathbb{R}^\sim$ is a field.
2. The ring $\mathbb{R}^\sim$ validates the axiom of micro-affinity and several related axioms.
3. There is a fully faithful functor $i : \mathrm{Mnf} \to \mathcal{E}$ embedding the category of smooth manifolds into $\mathcal{E}$.
4. The ring $\mathbb{R}^\sim$ coincides with $i(\mathbb{R}^1)$, the image of the real line in $\mathcal{E}$.

It is the culmination of a long line of research by several authors that several well-adapted models of synthetic differential geometry exist, see Moerdijk and Reyes (1991). By the conditions imposed in Definition 7, for any such topos $\mathcal{E}$ the following transfer principle holds: If $f, g : \mathbb{R} \to \mathbb{R}$ are smooth functions, then $f' = g$ (in the ordinary sense of the derivative) if and only if $i(f)' = i(g)$ in $\mathcal{E}$ (in the synthetic sense of the derivative).

Hence the nilsquare infinitesimal numbers of synthetic differential geometry may freely be employed as a convenient fiction when computing derivatives. Because the theorem on the existence of well-adapted models has a constructive proof, any proof making use of these infinitesimals may mechanically be unwound to a (longer and more complex) proof which only refers to the ordinary reals.

4.4.5 On the Importance of Language

The verification of the field property of $\mathbb{R}^\sim$ in Sect. 4.4.3 on page 90 demonstrates a basic feature of the Kripke–Joyal translation rules: The translation "$\mathcal{E} \models \varphi$" of a statement φ is usually quite complex, even if φ is reasonably transparent.

The language of toposes derives its usefulness for mathematical practice from this complexity reduction: In some cases, the easiest way to prove a result (about objects of the standard topos) is

1. to observe that the claim is equivalent to the Kripke–Joyal translation of a different (typically more transparent) claim about objects of some problem-specific relevant topos and then
2. to verify this different claim, reasoning internally to the topos.

One can always mechanically eliminate the topos machinery from such a proof, by translating all intermediate statements following the Kripke–Joyal translation rules and unwinding the constructive soundness proof of Theorem 1. This unwinding typically turns transparent internal proofs into complex external proofs – proofs which one might not have found without the problem-adapted internal language provided by a custom-tailored topos.

Acknowledgments We are grateful to Andrej Bauer, Martin Brandenburg, Sina Hazratpour, Matthias Hutzler, Marc Nieper-Wißkirchen and Alexander Oldenziel for invaluable discussions shaping this work, to Moritz Laudahn, Matthias Hutzler and Milan Zerbin for their careful criticism of earlier drafts and to Todd Lehman for the code for parts of Fig. 4.1. This note also profited substantially from the thorough review by three anonymous referees, whose efforts are much appreciated and gladly acknowledged. We thank the organizers and all participants of the Mussomeli conference of the Italian Network for the Philosophy in Mathematics, where this work was presented, for creating an exceptionally beautiful meeting. In particular, we are grateful to Neil Barton, Danielle Macbeth, Gianluigi Oliveri and Lorenzo Rossi for valuable comments.

References

Artin, M., A. Grothendieck, and J. Verdier. 1972. *Théorie des topos et cohomologie étale des schémas (SGA 4)*, Lecture Notes in Mathematics, vols. 269, 270, 305. Berlin: Springer.

Awodey, S., C. Butz, A. Simpson, and T. Streicher. 2014. Relating first-order set theories, toposes and categories of classes. *Annals of Pure and Applied Logic* 165: 428–502.

Barton, N., and S.-D. Friedman. 2019. Set theory and structures. In *Reflections on the Foundations of Mathematics*, Synthese Library, vol. 407, ed. S. Centrone, D. Kant, and D. Sarikaya, 223–253. Springer.

Bauer, A. 2005. Realizability as the connection between computable and constructive mathematics. http://math.andrej.com/asset/data/c2c.pdf.

Bauer, A. 2012. Intuitionistic mathematics and realizability in the physical world. In *A Computable Universe*, ed. H. Zenil. Singapore: World Scientific Pub Co.

Bauer, A. 2013. Five stages of accepting constructive mathematics. Lecture at the Institute for Advanced Study, https://video.ias.edu/members/1213/0318-AndrejBauer.

Bauer, A. 2015. An injection from the baire space to natural numbers. *Mathematical Structures in Computer Science* 25(7): 1484–1489.

Bishop, E., and D. Bridges. 1985. *Constructive Analysis*. Berlin: Springer.

Blechschmidt, I. 2017. *Using the internal language of toposes in algebraic geometry*. Ph. D. thesis, University of Augsburg. https://rawgit.com/iblech/internal-methods/master/notes.pdf.

Blechschmidt, I. 2018. An elementary and constructive proof of Grothendieck's generic freeness lemma. https://arxiv.org/abs/1807.01231.

Blechschmidt, I. 2020. A general nullstellensatz for generalized spaces. https://rawgit.com/iblech/internal-methods/master/paper-qcoh.pdf.

Borceux, F. 1994. *Handbook of Categorical Algebra: Volume 3, Sheaf Theory*, Encyclopedia of Mathematics and Its Applications. Cambridge: Cambridge University Press.

Butterfield, J., J. Hamilton, and C. Isham. 1998. A topos perspective on the Kochen–Specker theorem, I. Quantum states as generalized valuations. *International Journal of Theoretical Physics* 37(11): 2669–2733.

Caramello, O. 2014. Topos-theoretic background. https://www.oliviacaramello.com/Unification/ToposTheoreticPreliminariesOliviaCaramello.pdf.

Caramello, O. 2018. *Theories, Sites, Toposes: Relating and Studying Mathematical Theories Through Topos-Theoretic 'Bridges'*. Oxford: Oxford University Press.

Ceĭtin, G. 1962. Algorithmic operators in constructive metric spaces. *Tr. Mat. Inst. Steklova* 67: 295–361.

Couture, J., and J. Lambek. 1991. Philosophical reflections on the foundations of mathematics. *Erkenntnis* 34: 187–209.

Crosilla, L. 2018. Exploring predicativity. In *Proof and Computation*, ed. K. Mainzer, P. Schuster, and H. Schwichtenberg, 83–108. Singapore: World Scientific.

Escardó, M. 2013. Infinite sets that satisfy the principle of omniscience in any variety of constructive mathematics. *The Journal of Symbolic Logic* 78(3): 764–784.

Fabert, O. 2015a. Floer theory for Hamiltonian pde using model theory. https://arxiv.org/abs/1507.00482.

Fabert, O. 2015b. Infinite-dimensional symplectic non-squeezing using non-standard analysis. https://arxiv.org/abs/1501.05905.

Fourman, M. 1980. Sheaf models for set theory. *Journal of Pure and Applied Algebra* 19: 91–101.

Goldblatt, R. 1984. *Topoi: The Categorical Analysis of Logic*, Studies in Logic and the Foundations of Mathematics, vol. 98. Amsterdam: Elsevier.

Hakim, M. 1972. *Topos annelés et schémas relatifs*, Ergeb. Math. Grenzgeb., vol. 64. Berlin: Springer.

Hamkins, J. 2012. The set-theoretic multiverse. *Review of Symbolic Logic* 5: 416–449.

Hamkins, J., and A. Lewis. 2000. Infinite time turing machines. *The Journal of Symbolic Logic* 65(2): 567–604.

Heunen, C., N. Landsman, and B. Spitters. 2009. A topos for algebraic quantum theory. *Communications in Mathematical Physics* 291(1): 63–110.

Hyland, M. 1982. The effective topos. In *The L. E. J. Brouwer Centenary Symposium*, ed. A.S. Troelstra and D. van Dalen, 165–216. Amsterdam: North-Holland.

Johnstone, P.T. 2002. *Sketches of an Elephant: A Topos Theory Compendium*. Oxford: Oxford University Press.

Kock, A. 2006. *Synthetic Differential Geometry*, London Mathematical Society Lecture Note Series, vol. 333, 2nd ed. Cambridge: Cambridge University Press. http://home.math.au.dk/kock/sdg99.pdf.

Kock, A. 2020. New methods for old spaces: Synthetic differential geometry. In *New Spaces in Mathematics and Physics*, vol. 1, ed. M. Anel and G. Cartren. Cambridge: Cambridge University Press. https://arxiv.org/abs/1610.00286.

Kreisel, G., D. Lacombe, and J. Shoenfield. 1959. Partial recursive functionals and effective operations. In *Constructivity in Mathematics: Proceedings of the Colloquium Held in Amsterdam, 1957*, ed. A. Heyting, 290–297. Amsterdam: North-Holland.

Lambek, J. 1994. Are the traditional philosophies of mathematics really incompatible? *The Mathematical Intelligencer* 16(1): 56–62.

Leinster, T. 2011. An informal introduction to topos theory. *Publications of the nLab* 1(1).

Lombardi, H., and C. Quitté. 2015. *Commutative Algebra: Constructive Methods*. Netherlands: Springer.

Longley, J., and D. Normann. 2015. *Higher-Order Computability*, Theory and Applications of Computability. Berlin: Springer.

Mac Lane, S., and I. Moerdijk. 1992. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*, Universitext. New York: Springer.

Maietti, M. 2010. Joyal's arithmetic universes as list-arithmetic pretoposes. *Theory and Applications of Categories* 23(3): 39–83.

Maietti, M., and S. Vickers. 2012. An induction principle for consequence in arithmetic universes. *Journal of Pure and Applied Algebra* 216(8–9): 2049–2067.

Marquis, J.-P. 2013. Categorical foundations of mathematics. *The Review of Symbolic Logic* 6(1): 51–75.

Marquis, J.-P. 2019. Category theory. In *The Stanford Encyclopedia of Philosophy*, ed. E. Zalta, Fall 2019 ed. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2019/entries/category-theory/.

McLarty, C. 1990. The uses and abuses of the history of topos theory. *The British Journal for the Philosophy of Science* 41(3): 351–375.

McLarty, C. 2004. Exploring categorical structuralism. *Philosophy of Mathematics* 12(1): 37–53.

Melikhov, S. 2015. Mathematical semantics of intuitionistic logic. https://arxiv.org/abs/1504.03380.

Menni, M. 2000. *Exact completions and toposes*. Ph. D. thesis, University of Edinburgh. https://www.lfcs.inf.ed.ac.uk/reports/00/ECS-LFCS-00-424/ECS-LFCS-00-424.pdf.

Milne, J. 2013. Lectures on Étale cohomology. https://www.jmilne.org/math/CourseNotes/LEC.pdf.

Mines, R., F. Richman, and W. Ruitenburg. 1988. *A Course in Constructive Algebra*, Universitext. New York: Springer.

Moerdijk, I., and G. Reyes. 1991. *Models for Smooth Infinitesimal Analysis*. Springer. New York.

Phoa, W. 1992. An introduction to fibrations, topos theory, the effective topos and modest sets. Technical report, University of Edinburgh. http://www.lfcs.inf.ed.ac.uk/reports/92/ECS-LFCS-92-208/.

Pospíšil, B. 1937. Remark on bicompact spaces. *Annals of Mathematics* 38(4): 845–846.

Shulman, M. 2010. Stack semantics and the comparison of material and structural set theories. https://arxiv.org/abs/1004.3802.

Shulman, M. 2016. Categorical logic from a categorical point of view (draft for aarms summer school 2016). https://mikeshulman.github.io/catlog/catlog.pdf.

Streicher, T. 2004. Introduction to category theory and categorical logic. https://www.mathematik.tu-darmstadt.de/~streicher/CTCL.pdf.

Streicher, T. 2009. Forcing for IZF in sheaf toposes. *Georgian Mathematical Journal* 16: 203–209.

Tao, T. 2012. A cheap version of nonstandard analysis. Blog post, https://terrytao.wordpress.com/2012/04/02/a-cheap-version-of-nonstandard-analysis/.

Tennenbaum, S. 1959. Non-archimedean models for arithmetic. *Notices of the American Mathematical Society* 6: 270.

van den Berg, B., and J. van Oosten. 2011. Arithmetic is categorical. https://www.staff.science.uu.nl/~ooste110/realizability/arithcat.pdf.

van Oosten, J. 2008. *Realizability: An Introduction to Its Categorical Side*, Studies in Logic and the Foundations of Mathematics, vol. 152. Amsterdam: Elsevier.

Vickers, S. 2016. Sketches for arithmetic universes. https://arxiv.org/abs/1608.01559.

Chapter 5
Rescuing Implicit Definition from Abstractionism

Daniel Waxman

Abstract Neo-Fregeans in the philosophy of mathematics hold that the key to a correct understanding of mathematics is the *implicit definition* of mathematical terms. In this paper, I discuss and advocate the rejection of abstractionism, the putative constraint (latent within the recent neo-Fregean tradition) according to which all acceptable implicit definitions take the form of abstraction principles. I argue that there is reason to think that neo-Fregean aims would be better served by construing the *axioms of mathematical theories* themselves as implicit definitions, and consider and respond to several lines of objection to this thought.

Keywords Abstractionism · Neo-Fregeanism · Logicism · Hume's Principle

5.1 Introduction

Neo-Fregeans in the philosophy of mathematics hold that the key to a correct understanding of mathematics is the *implicit definition* of mathematical terms.[1] Such implicit definitions, they believe, play two roles. First, they play the *semantic* role of introducing terms which—according to a battery of background semantic and metaphysical views—successfully refer to mathematical objects. Second, they play the *epistemic* role of allowing for a priori justification or knowledge of at least basic propositions concerning the objects whose reference has been secured, and thereby (via plausibly a priori logical resources) allow for justification or knowledge of a substantial range of non-basic propositions too. If neo-Fregeans are right,

[1] Many have contributed to the neo-Fregean programme in one way or another. See Wright (1983) for the *locus classicus*, Hale and Wright (2000) for their conception of implicit definition and the role it plays, and the essays in Hale and Wright (2001).

D. Waxman (✉)
Department of Philosophy, National University of Singapore, Singapore

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_5

implicit definition holds out the prospect of providing a fully satisfying solution to the famous challenge raised by Benacerraf (1973), who saw the philosophy of mathematics pulled in mutually incompatible directions by the desire, on the one hand, to give mathematical sentences a face value reading and the need, on the other, to explain how a tractable epistemology of mathematical belief is possible.

Consider arithmetic, around which much of the discussion has revolved. According to neo-Fregeans, arithmetic is grounded (in a sense to be explained) in the stipulation of what has come to be known as Hume's Principle:

(HP) $\forall F \forall G\,(\#F = \#G \leftrightarrow F \approx G)$

where $\approx$ abbreviates the claim that there is a bijection between F and G (which can be defined in second-order logic).

The idea is that HP is to be viewed as an implicit definition of the cardinality operator "the number of…", denoted by #. Much neo-Fregean effort has been expended on developing a general account of implicit definition intended to vindicate the claim that, in the best cases—of which HP is supposed to be a representative example—implicit definitions can perform a great deal of valuable philosophical work. Here are some of the main benefits that have been claimed for HP, with directly analogous virtues carrying over to the other 'good cases' of implicit definition:

1. HP, when understood as an implicit definition, bestows a clear sense on the (previously undefined) "number of" operator "#". This definition serves to thereby introduce a range of complex singular terms of the form "#Φ", read as "the number of Φs".
2. Granted the success of HP as an implicit definition (and the truth of relevant instances of its right-hand-side), singular terms of the newly introduced form "#Φ" are guaranteed to refer. There is no *further* possibility of reference failure, that somehow there is no object to which "#Φ" refers, above and beyond the possibility of failure of the implicit definition itself.
3. HP is supposed to introduce a concept of a distinctive sortal kind—*cardinal number*—under which the referents of singular terms of the newly introduced form fall, and it is supposed to *explain* this newly introduced sortal kind. The explanation of the sortal kind succeeds, in particular, in introducing a new category of objects in such a way that explains why cross-categorical identifications (e.g. the claim that Julius Caesar is identical with the number 2) are unproblematically false.
4. Finally, HP is supposed to realize certain considerable epistemic benefits. Most important, it promises to sustain what Hale and Wright call the "traditional connection" between implicit definitions and a priori knowledge: as they put it, "to know both that a meaning is indeed determined by an implicit definition, and what meaning it is, ought to suffice for a priori knowledge of the proposition thereby expressed."[2] If this is right, HP is available as an item of a priori

[2] Hale and Wright (2000, 296).

> knowledge: it can be known, with no or at least minimal collateral epistemic work, simply via competent stipulation. It is well-known that HP, against a background of second-order logic, interprets second-order Peano Arithmetic. So perhaps second-order arithmetic in its entirety can come to be known a priori as well.[3]

Clearly, then, the benefits of implicit definitions—if the philosophical work in defending theses (1)–(4) can be pulled off—are vast. But although these benefits attach, allegedly, to implicit definitions in general, neo-Fregean attention has concentrated—I think it's fair to say exclusively—on implicit definitions of the following form:

$$\forall\alpha\forall\beta\,(\S\alpha = \S\beta \leftrightarrow \alpha \approx \beta)$$

where § is a term-forming operator, α and β are expressions of a certain type, and $\approx$ denotes an equivalence relation holding between entities of the relevant type. Call principles of that form *abstraction principles*. This paper will discuss, and ultimately advocate the rejection of, the view I will henceforth call *abstractionism*: that, in the good cases, abstraction principles enjoy certain metaphysical, semantic, and epistemic benefits not shared by putative implicit definitions of other forms. Building on a suggestion by John MacFarlane, my aim will be to motivate and explore the—in my view, considerable—attractions of a position that combines the neo-Fregean friendliness towards implicit definition with a broader non-abstractionism.[4] In particular, I will advocate a so-called *Hilbertian Strategy*, according to which *the axioms of mathematical theories* themselves constitute implicit definitions of the distinctive vocabulary used in their statement.

I will defend two main theses. The first is a *Parity Thesis*: if the benefits claimed for implicit definition by the neo-Fregeans are genuinely available for the taking, those benefits are equally available to proponents of the Hilbertian Strategy. My second thesis is that the Hilbertian Strategy in fact has significant advantages over the traditional approach. Unlike abstractionism, it generalizes seamlessly to the whole of mathematics. Furthermore, it opens up a plausible response to the Bad Company objection that has notoriously plagued the neo-Fregean program.

In a sense, the project can be viewed as an attempt to (perhaps subversively) appropriate the resources of implicit definition that have been defended so ardently and so arduously by the neo-Fregeans, and place them in the service of what might be called a neo-*Hilbertian* view.[5] Perhaps I should say "*re*-appropriate", for it is an irony of history that the self-described proponents of a distinctively neo-Fregean approach to the philosophy of mathematics are the chief contemporary

[3] The transition from "a theory within which the axioms of PA can be interpreted" to "arithmetic" should not go unnoticed, as Heck (2000) has emphasized, but neo-Fregeans plausibly hold that this transition is warranted.

[4] MacFarlane (2009). See Sect. 5.2 for more on MacFarlane's contribution.

[5] This term is aptly used by Hale and Wright (2009a) as a description of a view they oppose.

defenders of implicit definition as a means of gaining mathematical knowledge, given the antipathy that Frege himself manifested towards the notion both in his correspondence with Hilbert and in his post-*Grundlagen* writings.[6]

The plan is as follows. In Sect. 5.2, I briefly recap the constraints that Hale and Wright place on legitimate implicit definitions, and present (largely following MacFarlane 2009) a *prima facie* case that the Hilbertian Strategy satisfies each of them. In Sect. 5.3, I outline what I take to be the major advantages of such an approach, both in comparison to the orthdox neo-Fregean project and in its own right. I then turn in Sect. 5.4 to defend the view against a series of objections, the most serious of which are raised by Hale and Wright. These objections come in three varieties: semantic, epistemic, and concerns regarding the applicability of mathematics. I conclude that neo-Hilbertianism should, at the very least, be considered a serious rival to abstractionism.

5.2 Implicit Definition and the Hilbertian Strategy

5.2.1 Implicit Definition

The conception of implicit definition we'll consider is one elaborated by Hale and Wright in their essay *Implicit Definition and the A Priori*, although it was arguably latent in many earlier neo-Fregean writings. Hale and Wright (2000, 286) state the central idea as follows:

> we may fix the meaning of an expression by imposing some form of constraint on the use of longer expressions—typically, whole sentences—containing it.

More specifically [288]:

> We take some sentence containing—in the simplest case—just one hitherto unexplained expression. We stipulate that this sentence is to count as true. The effect is somehow to bring it about that the unexplained expression acquires a meaning of such a kind that a true thought is indeed expressed by the sentence—a thought which we understand and moreover know to be true, without incurring any further epistemological responsibility, just in virtue of the stipulation.

So implicit definition works by taking a sentence or sentences containing novel vocabulary to be *stipulatively true*. In the best cases, the idea runs, this will bestow the novel vocabulary with a meaning that serves to vindicate the stipulation. To return to the example of arithmetic, HP contains the novel operator "#" which is supposed to receive its meaning from the stipulation that, for all F and G, $\#F = \#G$ whenever F and G can be put into one-one correspondence.

Although it is beyond the scope of this paper to defend the merits of implicit definition in detail, it is worth saying something about the general theoretical

[6] For the correspondence, see Frege (1982).

orientation from which it arises. The view may seem alien to those who conceive of expressions having a meaning by somehow "latching on" to meaning-entities independently existing in some Fregean third realm. In contrast, the doctrine is seen naturally as arising from the combination of two commitments: (i) a generally use-based metasemantic account of how *sentential* semantic properties (in particular, truth) are fixed; and (ii) the explanatory priority of syntactic over semantic sub-sentential properties ("syntactic priority").[7]

In extremely compressed detail: the putatively mysterious idea of our "stipulating the truth" of a certain sentence is best understood as, roughly, commitment to employing that sentence in certain patterns of usage. The thought is that, against the background of a use-based metasemantic account, this will be possible to do in a way that ensures that such sentences are true. (It is helpful to compare the way in which inferentialists about logical connectives explain the semantic content of logical vocabulary). Syntactic priority is a complex package of views: in particular, it endorses three crucial moves: (i) from an expression's exhibiting the typical syntactic/inferential behaviour of a singular term to its being a *genuine* singular term; (ii) from an expression's being a genuine singular term figuring in a true sentence to its being a genuinely *referring* singular term; and (iii) from the fact that a singular term refers, to there being an *object* (as opposed to something of a different ontological category, e.g. a Fregean concept) to which it refers.[8]

Naturally, we can only hope to scratch the surface of issues concerning metasemantics, implicit definition, the syntactic priority thesis, and the connection between implicit definition and a priori justification here. My goal is not to defend the neo-Fregean cluster of views, but rather to argue for the conditional: *if* something in the ballpark of Hale and Wright's account of implicit definition is correct—i.e. in the best cases, implicit definitions can both (i) ensure that previously unmeaningful vocabulary can receive a meaning by appropriately figuring in sentences that are stipulated to be true, and (ii) provide a means for us to *know* the definitive sentence—then we have no reason to believe that the class of successful implicit definitions is restricted to abstraction principles.

What does "in the best cases" mean? Hale and Wright (2000) give five criteria they take to be individually necessary (and, tentatively, jointly sufficient) for an implicit definition to succeed:

Consistency The sentence serving as the vehicle of the definition must be *consistent*.

Conservativeness The sentence serving as the vehicle of the definition must be *conservative* over the base theory. Roughly, the extended theory ought not introduce new commitments concerning the ontology of the base theory.

[7] Much more on these views can be found in Hale and Wright (2001). See also Hale and Wright (2009b) on neo-Fregean metaontology and MacBride (2003) for a helpful overview of the semantic and metasemantic commitments of neo-Fregeanism.

[8] For an account of singular terms, see Hale's Chapters 1 and 2 of Hale and Wright (2001).

Less roughly, say that a theory T_2 is a *conservative* extension of T_1 if for any sentence ϕ in the language of T_1, if $\mathsf{T}_1 + \mathsf{T}_2 \vdash \phi$ then $\mathsf{T}_1 \vdash \phi$. Conservativeness is too strong of a condition to place on implicit definitions, at least if neo-Fregean proposals are to have a chance of getting off the ground. The reason is that Hume's Principle has consequences, e.g. that there are infinitely many objects, that may be expressible in the base language and yet not follow from the base theory. To get around this issue, neo-Fregeans have moved to what has become known as *Field-conservativeness*, which intuitively says that the extended theory has no new consequences *for the objects spoken of by the base theory.*[9] To express this formally, we need some notation. Let $P(x)$ be a predicate that doesn't occur in T_1 or T_2 (intended to pick out all and only the "old" objects spoken of by T_1). For each sentence ϕ in the language of T_1, let ϕ^* be the result of relativizing all quantifiers occurring in ϕ to $P(x)$ and let T_1^* be the theory whose axioms are the sentences ψ^*, where ψ is an axiom of T_1.[10]

This allows the relevant notion of conservativeness to be defined: T_2 is a *Field-conservative extension* of T_1 if, for any sentence ϕ in the language of T_1, if $\mathsf{T}_1^* + \mathsf{T}_2 \vdash \phi^*$, then $\mathsf{T}_1 \vdash \phi$.

The proposal, then, is that implicit definitions must be Field-conservative over the theory to which they are added.

Harmony In Wright and Hale's discussion, Harmony is a constraint that applies primarily to implicit definitions of logical expressions (in particular, expressions with both an "introduction rule" and an "elimination rule"), intended to rule out unharmonious connectives like Pryor's "tonk" and its dual. It is unclear that this constraint has any bearing on the mathematical cases we are concerned with here, so I will pass over it without further comment.[11]

Generality This condition holds that the meaning that new terms receive from implicit definitions must satisfy Gareth Evans (1982)'s *Generality Constraint:* very roughly, if new sub-sentential expressions are introduced by a definition, one who grasps the definition must thereby be provided with the means of understanding the meaning of arbitrary sentences composed of (grammatically appropriate) combination of the new expression and antecedently understood vocabulary. This is obviously connected with the Caesar problem: if, as neo-Fregeans contend, HP construed as an implicit definition is able to provide singular terms such as $\#F$ with a meaning, then the Generality Constraint requires (as a special case) that

[9] See Field (2016, 11) for this notion of conservativeness, and Hale and Wright (2001, 297) for the thought that it is the relevant notion in stating this constraint on implicit definition.

[10] I assume that the axioms of theories are *sentences* rather than open formulae.

[11] See for instance Tennant (1978) and Dummett (1991b) for discussions of Harmony in the logical setting. More recently, Wright (2016) suggests that Hume's Principle is best understood as functioning more like a rule of inference than an axiom. A full discussion of Harmony in the context of non-logical rules, however, would take us too far afield.

it also provide us with a means of understanding mixed identity sentences like $\#F =$ Julius Caesar.

Anti-arrogance By contrast with the previous—logical and semantic—conditions, Anti-arrogance is an epistemic constraint. Hale and Wright put it as follows. An implicit definition is arrogant if:

> the truth of the vehicle of the stipulation is hostage to the obtaining of conditions of which it's reasonable to demand an independent assurance, so that the stipulation cannot justifiably be made in a spirit of confidence, "for free"

Here is a proposal to make this more precise. Say that a sentence (understood as the vehicle of an implicit definition) S is arrogant if (i) there is some condition C that must be satisfied for the truth of S; (ii) we require justification that C is satisfied in order to have justification in S; and (iii) we have no justification that C is satisfied.[12]

There are two points worth noting. One is that condition (ii) is somewhat schematic: to flesh it out, we need to know which conditions require antecedent justification. This involves deep questions in epistemology, which cannot fully be discussed here. But there are different possible views here, ranging from the very conservative (we must have justification that *all* conditions necessary for the truth of the stipulation are satisfied) to the very liberal (we do not need justification that *any* are). Hale and Wright are not very explicit about where on this spectrum they lie, but their discussion seems to situate them towards the more conservative end. I will follow them in this respect. Dialectically, this is appropriate: the more conservative the approach, the fewer legitimate implicit definitions there will be, so proceeding this way does not give the Hilbertian unfair advantage.[13] The second point to note is that, due to condition (iii), whether an implicit definition is arrogant in the above sense will be sensitive to one's epistemic position. In other words, it is possible for a stipulation to be arrogant in the hands of one epistemic agent, and not in the hands of another, depending on which conditions they have justification to accept. This will prove later to be of significance.[14]

---

[12] For the sake of clarity, I understand justification in the propositional sense: roughly, what it would be appropriate for someone in one's epistemic position to believe, whether or not one in fact has the belief in question. I mean to be as neutral as possible here, and in particular not to exclude varieties of entitlement, i.e. "default" or "non-earned" forms of justification. See Wright (2016), and Hale and Wright (2001, 127) for evidence that this is how they understand our justification in the preconditions for the stipulation of HP.

[13] Ebert and Shapiro (2009), Section 6, discuss some of the options one might adopt here, and argue that neither (what I have called) the very conservative nor the very liberal approaches are very attractive. However, their discussion of the conservative approach seems to me to be marred by a conflation of the *knowability* or *justifiability* of consistency with its *provability*, which they rightly take to be ruled out for Godelian reasons. I discuss this, and the epistemology of consistency more generally, in other papers.

[14] Anti-arrogance ought to be distinguished from other conditions that might be confused with it. One is *conditionality*; it is not equivalent to saying that implicit definitions have a conditional as their main connective. (To see that it is not sufficient, consider any arrogant stipulation A and consider $\top \rightarrow A$, where $\top$ is some logical truth. To see that it is not necessary, take Hume's

At this stage, it is time to introduce the proposed alternative to the neo-Fregean strategy and to evaluate it against these criteria.

5.2.2 The Hilbertian Strategy

Return again to the case of arithmetic. Neo-Fregean abstractionists want to explain our understanding of arithmetic basically as follows: we begin by stipulating Hume's Principle as a definition of "the number of…", and then we use (second-order) logical resources to derive certain theorems, including, most importantly, the axioms of PA. This result is now known as Frege's theorem, and is surely one of Frege's most substantial mathematical achievements.[15] According to neo-Fregeans, Frege's Theorem provides a means of recovering the PA axioms and thereby the whole of arithmetic in an epistemically responsible way, since, they claim, (i) Hume's Principle is something that we can come to know a priori, and (ii) knowledge can be transmitted via competently carried-out second-order logical deductions.[16]

But why the need to go via HP? What, exactly, is mandatory about taking the neo-Fregean route to the Peano Axioms, as opposed to the following procedure? Lay down the (conjunction of) the Peano Axioms as a stipulation, intended to implicitly define the expression *S*—denoting the successor function—the numerical singular terms—0, 1, 2, etc—and a predicate $\mathbb{N}$ applying to all and only natural numbers. Instead of using HP to implicitly define the notion of natural number and to serve as the premise for a derivation of the PA axioms, the proposal is that *the axioms themselves* are understood as the definitive principles introducing that notion. This is what I will call the Hilbertian Strategy applied to arithmetic. More generally, it takes the axioms of some target theory as themselves serving as an implicit definition of the central notions involved.

It should be simple to see how the Hilbertian Strategy can, in principle, be extended to any other axiomatizable mathematical theory. To take just a few salient examples, the axiomatic theory of the real numbers (axiomatized e.g. as a complete ordered field), the complex numbers (axiomatized as e.g. the algebraic closure of the reals, or as a field of characteristic 0 with transcendence degree $\mathfrak{c}$ over $\mathbb{Q}$), set

Principle itself). Anti-arrogance is also importantly distinct from Conservativeness. Consider Goldbach's conjecture which (let us suppose) is a truth of arithmetic for which we lack a proof and, consequently, justification. Now consider the stipulation of the theory: PA + Goldbach's conjecture. This is, *ex hypothesi*, a conservative extension, for Goldbach's conjecture follows from PA. But it is nevertheless arrogant, for we plausibly need *independent* assurance that Goldbach's conjecture is true before we can have any justification that the theory is true, i.e. that there are entities that simultaneously satisfy PA and Goldbach's conjecture.

[15] For details of the modern rediscovery of Frege's Theorem and the contributions (in various parts) of Geach, Parsons, Tennant, Wright, Boolos, and Heck, see Burgess (2005, Ch 3).

[16] In doing so they presumably appeal to a plausible epistemic closure principle.

theory (in any of its usual axiomatizations, e.g. ZFC, NBG, etc). In general there is a version of the Hilbertian Strategy available for any axiomatic theory: one merely takes the axioms themselves as an implicit definition of the predicates, relations, and constants involved in the axiomatization.[17]

Orthodox neo-Fregeans will no doubt be filled with the conviction that whatever the apparent attractions of the Hilbertian Strategy, they are the result of theft over honest toil. But they must face the question, first raised by John MacFarlane: what strictures on implicit definition would this procedure violate, in the case of arithmetic in particular and for other mathematical theories more generally? Let us take the conditions in turn.[18] I will attempt to make a *prima facie* case that the two approaches are, at the very least, on a par. Later, in Sect. 5.4, I will consider and reply to various subtle arguments to the effect that, despite first appearances, neo-Fregeans are in a better position than Hilbertians to show that the relevant criteria are satisfied.

Consistency If HP is consistent, then so is PA. This is because PA is relatively interpretably within HP. On minimal epistemic assumptions this implies that we have at least as much reason for believing that PA is consistent as we do for believing that HP is consistent. So if the consistency constraint is satisfied by the neo-Fregean stipulation, it is satisfied by the Hilbertian stipulation.

Conservativeness The justification here is similar to that of consistency. If PA has any non-conservative implications over the base theory to which it is added, then so does HP, in virtue of the relative interpretability of PA within HP.

Harmony As mentioned, Harmony is irrelevant outside of the context of logical connectives.

Generality *Prima facie*, a stipulation of PA appears to be no better or worse off in regard to its ability to satisfy the Generality Constraint than does a stipulation of HP. It is true that a stipulation of PA does not, at least *prima facie*, fix the meaning or truth-conditions of certain grammatically appropriate sentences involving the newly introduced terms, such as "$2 =$ Julius Caesar" or "$\mathbb{N}$(Caesar)" or "S(Caesar) $=$ (Augustus)". This is just to say that the Hilbertian faces a version of the Julius Caesar problem. However, the Caesar problem is notoriously pressing for abstractionists too – indeed, it was precisely Frege's own reason for rejecting

[17] If the theory is finitely axiomatized, the definition can be given as a single conjunctive sentence. If it is schematically axiomatized, there are different options available: to use a truth predicate of some kind; to appeal to a device of infinite conjunction or some other kind of higher order resources; or to take the instances of the schema as collectively constituting the definition.

[18] Of course, this point is merely ad hominem against Hale and Wright in the absence of arguments that HP and PA *do* satisfy the requirements. I take it that the discussion to come addresses at least some worries about generality and arrogance; I discuss issues of consistency and conservativeness further in other work. The points in the remainder of this section largely follow MacFarlane (2009), who first pointed out that something like the Hilbertian Strategy appears to satisfy Hale and Wright's criteria.

(what in this context can only anachronistically be called) the neo-Fregean strategy for grounding arithmetic in HP. So, again *prima facie*, the two approaches are on a par. Naturally, neo-Fregeans have said much in response to the Caesar problem. As I will argue in Sect. 5.4, many of the resources to which they appeal can be equally well appropriated by the Hilbertian.

Anti-arrogance Again *prima facie*, a stipulation of PA appears to have equally—or less—demanding conditions for its truth than a stipulation of HP. Any model of HP can be expanded to a model of PA. For instance, given classical logic, both theories entail the existence of infinitely many objects, so both exclude finite domains. Consequently, it is difficult to see how PA might be *more* arrogant than HP. Certainly, one might not be justified in, e.g., believing that the universe co-operates in providing enough or the right kind of entities to allow the stipulations to be true. But if so, this is a reason to convict *both* implicit definitions of arrogance, not just PA. In Sect. 5.4, I consider and respond to some further subtle arguments from neo-Fregeans to the effect that the Hilbertian strategy is arrogant in a way that their own approach is not.

The point of the foregoing is that there is a strong *prima facie* case for the conditional claim: *if* the neo-Fregean Strategy with respect to arithmetic is successful, then so too is the acceptability of the Hilbertian Strategy. In fact, as we have seen, the Hilbertian Strategy appears if anything to be in a *better* position. Furthermore, since nothing particular to arithmetic was appealed to in these arguments, the reasoning above can be replicated in the general case: wherever the theory generated by an acceptable abstraction principle allows for interpretation of an axiomatic theory that we find of mathematical interest—that is just to say, whenever the neo-Fregean Strategy for explaining some portion of established mathematics is successful—then so too is the analogous instance of the Hilbertian Strategy. As I will now argue, the Hilbertian Strategy is not merely on a par with the neo-Fregean strategy: in fact, it has certain definite advantages.

5.3 Advantages of the Hilbertian Strategy

5.3.1 Abstractionism and Set Theory

A major thorn in the side of abstractionist neo-Fregeans has been an inability to develop a satisfactory theory of sets. As is well known, Frege's original approach, employing Basic Law V:

(BL V) $\forall F \forall G\,(\{x : Fx\} = \{x : Gx\} \leftrightarrow (\forall x(Fx \leftrightarrow Gx)))$

as the central abstraction principle governing the identity of sets, is inconsistent. The problem is simply this: no subsequent development of set theory has delivered a theory that can be plausibly considered foundational in the manner of ZFC, the

most popular and widely used set theory. The best-known approach, due to George Boolos, appeals to so-called New V:

(New V) $\forall F \forall G\,(\{x : Fx\} = \{x : Gx\} \leftrightarrow ((Bad(F)) \wedge (Bad(G)) \vee (\forall x\,(Fx \leftrightarrow Gx))))$

where $Bad(F)$ expresses the condition that F is "large", i.e. can be placed into bijection with the entire universe. The idea is to avoid paradox by encoding a class/set distinction into the implicit definition of the notion of set: set-sized concepts—those that are not equinumerous with the universe—form sets, while class-sized concepts do not. As Boolos showed, New V is consistent and able to obtain (with suitable definitions) many of the axioms of ZFC, including Extensionality, Empty-Set, Well-Foundedness, Pairing, Union, Separation, and Replacement. However, there is a clear sense in which this theory—call it Boolos Set Theory (BST)—lacks the ontological power of ZFC. To see this, note that BST is satisfied by the hereditarily finite sets: it follows that in it, the Axioms of Infinity and Powerset both fail, since neither hold in the hereditarily finite sets.[19] Given that the ontological power of ZFC is generated primarily by these two axioms, the resulting theory will be seen as admitting a severely impoverished ontology from the perspective of a believer in the universe of sets as standardly conceived. I do not know of any more promising abstractionist approaches to set theory.[20]

Of course, there are two reactions one might have to the seeming inability of abstractionism to recover a theory that is recognizably comparable to contemporary set theory in its scope and power. One can think, as Wright (2007, 174) tentatively advances, that if "it turns out that any epistemologically and technically well-founded abstractionist set theory falls way short of the ontological plenitude we have become accustomed to require, we should conclude that nothing in the nature of sets, as determined by their fundamental grounds of identity and distinctness, nor any uncontroversial features of other domains on which sets may be formed, underwrites a belief in the reality of that rich plenitude." But this is a radical conclusion indeed. Whatever one thinks of set theory, it is impossible to deny its importance as a foundational theory within contemporary mathematics, in effect

[19] The hereditarily finite sets are the sets with finite transitive closures: elements of V_κ for some finite κ, where this is the κth level of the von Neumann Universe.

[20] As Burgess (2004) has shown, by adding a reflection principle (and plural logical resources) to BST, all of the ZFC axioms are (remarkably) once again obtainable. It nevertheless ought to be clear that a theory formulated in this way will fail to satisfy abstractionist strictures, since to the best of my knowledge there is no way to straightforwardly write it as an abstraction principle. There is additionally a further problem that New V entails a *global* choice principle, and this may violate the conservativity requirement for implicit definitions (even the relevant and relatively weaker notion of Field-conservativeness as introduced above). See Shapiro and Weir (1999) for details. Fine (2002) has also worked extensively on the limits of abstraction principles; the theories he obtains, the nth order "general theory of abstraction" (for finite n), are in general, equi-interpretable with $n + 1$st order arithmetic. This is certainly enough to carry out much mathematics, but it is nevertheless a far cry from an adequate replacement for orthodox set theory. See Burgess (2005) for more on the limitations of Finean arithmetical theories.

providing the predominant ontological background and organizational framework within which mathematics is carried out. The fact that it cannot easily be seen obviously to follow from any abstraction principle governing the identity of sets is arguably more of a reason for rejecting that demanding constraint than it is for rejecting set theory itself.[21]

By contrast, compare how easily the proponent of the Hilbertian Strategy is able to respond to the challenge posed by contemporary set theory. It is unsurprisingly straightforward: merely take one's favourite set theory (say, ZFC, perhaps with some large cardinal axioms if one is feeling adventurous) and understand its axioms as implicitly defining the notions of set and membership. To be sure, for this to be a successful implicit definition of *set* and *membership*, ZFC must (like all putative implicit definitions) satisfy the previously mentioned conditions in order to succeed: ZFC must be consistent, conservative, not require any objectionably arrogant epistemic preconditions to be satisfied, etc. I do not wish to trivialize the question of whether these conditions hold. In my view, the question of whether our best theories, ZFC included, are consistent (and in particular how we manage to obtain justification to believe they are) is one of the most pressing and neglected in the philosophy of mathematics. But this is not the place to discuss the issue further; for our purposes it's enough to note that, naturally, any neo-Fregean abstraction principle purporting to recover set theory would necessarily face the same challenges.

In short: the neo-Fregean's prospects for recovering a theory capable of playing the foundational role of set theory is questionable, and there is reason to be optimistic about the Hilbertian's prospects for doing the same.

5.3.2 How to Rid Oneself of Bad Company

Another major issue for neo-Fregeans is what has come to be known as the Bad Company Objection. There are a great number of possible abstraction principles that one might be tempted to adopt; but—and here is the rub—some of these principles lead to unacceptable results. The most obvious kind of unacceptability is exemplified by Basic Law V, which is inconsistent. But there are other, less obvious

[21] In conversation, Crispin Wright has raised the interesting idea that an abstractionist treatment of set theory might have the benefit of providing a principled distinction between features of set theory that are (my word, not his) "intrinsic", flowing from the nature of sets—i.e. those that follow from the abstraction principle (whatever it may be) governing sets—and those that are "extrinsic". (For instance, if we go with Boolos Set Theory, Powerset and Infinity will count as extrinsic axioms in this sense.) While this is intriguing, I think (if it is to recapture a theory with anything like the strength of ZFC), it will end up raising more epistemological difficulties than it solves. In particular, the question of the justification of any extrinsic axioms will become urgent in the face of something like Benacerraf's original dilemma: and, of course, on an abstractionist view, the resources of implicit definition will be unavailable to play any epistemic role here.

kinds of unacceptability also. For instance, there are pairs of abstraction principles that are individually consistent (and conservative) which nevertheless, when taken together, result in inconsistency.[22] So the problem is essentially this: given that not all consistent and conservative abstraction principles are acceptable, some account needs to be provided of which principles *are* acceptable. The problem generalizes in the obvious way to implicit definitions in general (of which abstraction principles are particular examples). How might this problem be resolved?

Neo-Fregeans face a difficult technical challenge here: to investigate the logical and mathematical features of abstraction principles and hope that some criteria can be found to distinguish, in a principled way, between good and bad cases.

I don't pretend to be in a position to offer anything like this kind of solution. Nevertheless, I do want to argue that there's a good sense in which the Hilbertian Strategy allows us to sidestep these worries by showing how implicit definitions can be combined, as long as the theories in question are consistent, to yield in effect the whole of contemporary mathematics. Here's a rough idea of what I have in mind. The Appendix contains a proof of the following result:

Let T_m (for "mathematical") and T_b (for "background") be theories whose languages share no individual constants. As before, let T_m^* be the theory that results from relativizing the quantifiers of T_m to a fresh predicate, and let T_b^* be the theory that results from relativizing the quantifiers of T_b to a fresh (and different) predicate.

Theorem 1 *If T_m and T_b are consistent first-order theories, then $\mathsf{T}_b^* + \mathsf{T}_m^*$ is consistent and T_m^* is Field-conservative over T_b.*

The restriction to first-order theories is important: unfortunately, the result does not hold for second-order theories.[23] However, a slightly weaker result can be shown for second-order theories. First a definition:

Definition (Field*-conservativeness.) T_2 is a Field*-conservative extension of T_1 if, for any sentence ϕ in the language of T_1, if $\mathsf{T}_1^* + \mathsf{T}_2 \vDash \phi^*$, then $\mathsf{T}_1 \vDash \phi$.

where $\vDash$ is understood as *full* semantic consequence.[24]
Then we have:

Theorem 2 *If T_m and T_b are satisfiable second-order theories, then $\mathsf{T}_b^* + \mathsf{T}_m^*$ is satisfiable and T_m^* is Field*-conservative over T_b.*

[22] For more on Bad Company, see e.g. Wright (1999), Shapiro and Weir (1999), a special issue of *Synthese* edited by Linnebo (2009), and several papers in Cook (2016).

[23] Proof sketch: let T_b be $\mathsf{PA}_2 + Con_{\mathsf{PA}_2}$ and let T_m be $\mathsf{PA}_2 + \neg Con_{\mathsf{PA}_2}$, formulated in a disjoint language (i.e. with different arithmetical constants and predicate-symbols). Then each theory is individually consistent; but $\mathsf{T}_b^* + \mathsf{T}_m^*$ is inconsistent, since it violates Internal Categoricity in the sense of Button and Walsh (2018, Chapter 10).

[24] See Shapiro (1991) for a definition.

Take Theorem 1 first. I'd like to gloss this result as saying that, if you start with a consistent base theory, and add to that theory an arbitrarily chosen consistent mathematical theory then—after the theories are cleaned up in what I take to be a wholly philosophically defensible way—the resulting theory is going to be (i) consistent and (ii) Field-conservative over the base theory. The philosophical upshot is, I claim, that the Hilbertian Strategy can provide an operational solution to the Bad Company objection: the result shows that if we add a new mathematical theory obtained via the Hilbertian Strategy, then (as long as the theory we attempt to add is consistent, which as we have seen is a condition on its success as an implicit definition in the first place), the end result will be itself consistent and conservative over the base theory to which it is added. What is more, this is a strategy that can be extended indefinitely: as long as the new theory we add at each stage is consistent, then there is *never* any risk of ending up in inconsistency or with objectionably non-conservative consequences.[25] There is no prospect (as there is with abstraction principles in general) of falling into inconsistency by way of adding individually consistent principles/theories.

Three brief clarifications are in order. First, a word on the "cleaning up" of the theories. All I really mean by this is that before the theories are added together, the quantifiers of each are relativized to a predicate that is intended to pick out all and only the entities spoken of by the theory. Suppose we have a base theory that does not make any claims whatsoever about sets—a physical theory, say—and suppose we want implicitly to define a notion of set via the Hilbertian Strategy. The idea is that we introduce new predicates "set" and "non-set" and relativize all of the quantifiers of the relevant theories to those predicates. That way, we will not be saddled with set-theoretic claims like

$$\forall x \forall y \, (x = y \leftrightarrow \forall z \, (z \in x \leftrightarrow z \in y))$$

that have the (absurd) consequence of identifying all, e.g., physical objects that are non-sets. For in its relativized form

$$\forall x \forall y \, (Set(x) \wedge Set(y) \rightarrow (x = y \leftrightarrow \forall z \, (Sz \rightarrow (z \in x \leftrightarrow z \in y))))$$

the principle will be explicitly restricted only to sets. This move seems philosophically well-motivated for reasons independent of anything to do with the present project: theories, presumably, should always be written this way when we are being fully explicit, and this is especially true when our intention is to implicitly define

[25] This is true, at least, when we "only" extend our theories finitely many times; which, I submit, is more than enough in practice.

new predicates that are intended to range over objects to which our old language did not refer.[26,27]

Second, it's worth noting that matters are slightly more complex for second-order theories: the mere consistency of the theories in question does not suffice, and the result as stated uses a kind of generalization of Field-conservativeness that involves second-order semantic consequence. While this is certainly a qualification worth mentioning, it nevertheless seems that the results taken together give us all that we need. For first-order theories—including the foundational case of most interest (set theory as codified in ZFC)—consistency suffices. If the theories in question are second-order, then the stronger condition that they must be (individually) satisfiable—where satisfiability is, in effect, the semantic analogue of consistency—is needed. This is very much in the spirit of the consistency requirement: indeed, for first-order theories, consistency and satisfiability in the relevant sense are coextensive.

Third—and this is one reason why I do not claim to have a fully diagnostic solution to Bad Company worries in general—I concede that the result mentioned does not help much in illuminating the question of Bad Company for abstractionists, since abstraction principles cannot easily conform to the relativization procedure discussed above. That said, this last fact may be seen by some as an additional advantage of the Hilbertian Strategy over the neo-Fregean abstractionist alternative.

5.4 Objections and Replies

In this section I discuss and reply to a number of objections—in keeping with the spirit of the paper, primarily objections arising from a neo-Fregean perspective. These can be classified in three broad groups. On the conception considered here, implicit definition has both *semantic* and *epistemic* components. It is supposed to fix a meaning for the terms introduced, as well as to illuminate and explain our justification with respect to basic truths involving them. There are correspondingly two ways in which an implicit definition might be thought to be defective. First, it could be semantically defective and fail to establish a legitimate meaning for the terms it is intended to introduce. Second, even if it succeeds semantically, there

[26] For more see Field (2016, 12). NB: to say that mathematical theories should be *relativized* in this way (i.e. to make explicit which type of object they concern) is not to say that they should be written as *conditional* on the existence of objects of the relevant type. Thus what I say here does not take a stand on the dispute between Field (1984)—who argues that HP is only acceptable conditional on the claim that numbers exist—and Wright (in Chapter 6 of Hale and Wright 2001)—who replies that the concept of number, which of course on his view is given by HP itself, is required in order even to understand the antecedent of such a conditional. Thanks to a referee here.

[27] This relativization also allows the Hilbertian to sidestep worries about apparently incompatible theories: for instance, ZFC + CH vs ZFC + ¬CH: they will be relativized to different predicates (say, Set_1 and Set_2), avoiding any actual incompatibility.

may still be reasons why it fails to provide justification or knowledge. In addition, there is a third potential source of difficulty: whether Hilbertian theories are capable of *applications* of the sort we require from mathematics. In particular, there is a distinguished line of thought, arguably originating with Frege himself, according to which the applicability of mathematics must be placed at the center of matters; some have suggested that this is a reason to prefer the neo-Fregean Strategy. Each of these groups of complaint will be addressed in turn.

5.4.1 Objections to the Semantic Role of Hilbertian Definitions

5.4.1.1 Generality, Stipulation, and the Caesar Problem

Can instances of the Hilbertian Strategy really bestow a meaning on the novel vocabulary they contain? Hale and Wright think that there are serious doubts to be had. They ask us to consider the Ramsification of the conjunction of the axioms of PA: the sentence obtained by replacing the distinctively arithmetical vocabulary—S, $\mathbb{N}$, and 0—with variables, and existentially quantifying through these variables. Then, they consider the effect of stipulating this Ramsey sentence, and make three claims. First, it cannot be *meaning-conferring* in any plausible sense, for it contains no new vocabulary that could stand to receive a meaning. Secondly, although such a stipulation may (in a perverse way) introduce the concept ω-sequence, the stipulation of its truth is irrelevant in this respect: that concept could equally well have been introduced by saying that an ω-sequence is *whatever* satisfies the Ramsey sentence, without any claim that there are such things. Finally, they consider what the difference between a stipulation of the Ramsey sentence of PA, and PA itself would be. As they put it, a stipulation of the Ramsey sentence appears to amount to the command:

> Let there be an omega sequence!

But the stipulation of PA itself appears to amount to:

> Let there be an omega sequence whose first term is *zero*, whose every term has a unique *successor*, and all of whose terms are *natural numbers*![28]

so the complaint is:

> it is not clear whether there really is any extra content—whether anything genuinely additional is conveyed by the uses within the second injunction of the terms "zero", "successor" and "natural number". After all, in grasping the notion of an omega-sequence in the first place, a recipient will have grasped that there will be a unique first member, and a relation of succession. He learns nothing substantial by being told that, in the series whose existence has been stipulated, the first member is called "zero" and the relation of succession is called "successor"—since he does not, to all intents and purposes, know which are the objects for whose existence the stipulation is responsible. For the same reason, he

[28] Hale and Wright (2009a, 470).

> learns nothing by being told that these objects are collectively the "natural numbers", since he does not know what natural numbers are. Or if he does, it's no thanks to our stipulation.[29]

The objection can be reconstructed as follows:

(1) A stipulation of the Ramsey sentence for PA is not capable of playing a meaning-conferring role;
(2) A stipulation of the Ramsey sentence for PA is, in all relevant (i.e. meaning-conferring respects) equivalent to a stipulation of PA itself;
(3) So, a stipulation of PA is not capable of playing a meaning-conferring role.

This complaint raises subtle questions. To see why, it is helpful to ask the obvious question: why, if the present complaint is successful, do Hale and Wright not feel that it has any force against their own neo-Fregean view? Although they do not explicitly address the point in their discussion, my suspicion is that the answer is intimately related to the Julius Caesar problem; seeing why will allow us to assess the objection more adequately.

One way of understanding the Julius Caesar problem for neo-Fregeanism is as an accusation that HP is vulnerable to precisely the difficulty that is currently being alleged of PA: in particular, that HP is defective as a meaning-conferring definition, because it simply does nothing to tell us what the natural numbers are supposed to be. Although HP tells us *something* about the natural numbers and the "number of…" operator, namely that two concepts have the same number iff they can be put into bijection, it does nothing to enable to us to rule out claims like $2 =$ Julius Caesar, because HP is consistent with the natural numbers being anything whatsoever—even the familiar conqueror of Gaul himself. More generally, the complaint runs, HP puts us in no position to accept or reject mixed identity sentences of the form $\alpha = \beta$ where α is a canonical name for a number and where β is not; and this signals a major defect in it, understood as an implicit definition.

It is hard not to read Hale and Wright as adverting to this issue when they complain that a stipulation of PA does not put us in a position to "know which are the objects for whose existence the stipulation is responsible".[30] In other words, it seems they suspect that PA does nothing to single out the natural numbers: they could be anything, as far as we know from the definition, as long as there are enough of them and they have the relevant properties or stand in the relevant relations.

Hale and Wright, naturally, think that they have the resources to overcome the Caesar problem insofar as it threatens HP. In very compressed form, their solution is to appeal to the notion of a *pure sortal* predicate. Roughly, an entity falls under a pure sortal predicate if it is "a thing of a particular generic kind—a person, a tree, a river, a city or a number, for instance—such that it belongs to the essence of the object to be a thing of that kind."[31] For Hale and Wright, it is intimately part of the ideology of a pure sortal that it comes with an associated *criterion of identity*: a

[29] Hale and Wright (2009a, 471-2).
[30] Hale and Wright (2009a, 470).
[31] Hale and Wright (2001, 387).

principle that canonically determines the truth of identity-statements that contain terms referring to entities of the relevant sort.[32] For instance, Hume's Principle can be understood as a criterion of identity for numbers (i.e. numbers are equal when they apply to equinumerous concepts); the Axiom of Extensionality can be understood as a criterion of identity for sets (i.e. sets are equal when they have the same members); spatio-temporal continuity can be understood as a criterion of identity for physical objects across time; and so on. And here the possibility of an objection to the Hilbertian Strategy on behalf of the abstractionist opens up. For the instances of the Hilbertian Strategy that we have been considering do not—unlike instances of the abstractionist strategy—appear to come ready made with criteria of identity for the new sortal terms distinctively introduced.

This, I think, is among the strongest objections that the neo-Fregean is in a position to make. In short: abstraction principles function as criteria of identity; so, without abstraction principles, the Hilbertian does not have recourse to a criterion of identity for the objects characterized by the newly-defined vocabulary, and therefore lacks the resources to answer the Julius Caesar problem. What can be said in response?

My suggestion is first briefly to step back and examine the reasons why a need for a criterion of identity appears to arise in the first place. In their solution to the Caesar problem, Hale and Wright (2001, 389) present a picture in which:

> all objects belong to one or another of a smallish range of very general categories, each of these subdividing into its own respective more or less general pure sorts; and in which all objects have an essential nature given by the most specific pure sort to which they belong. Within a category, all distinctions between objects are accountable by reference to the criterion of identity distinctive of it, while across categories, objects are distinguished by just that—the fact that they belong to different categories.

The response I propose on behalf of the Hilbertian is to suggest that in order to sustain this picture—or at least, the part of it that involves *mathematical* categories—it is sufficient that we are provided with what we might call *theory-internal adjudications of identity*. The idea is that, for at least appropriately chosen mathematical theories, the theories themselves in some sense "tell us all we need to know" about the identity of the objects that the theories are about. Let me try and motivate this with some examples.

1. Set theory. It follows from the axioms of Zermelo Fraenkel set theory that

$$\forall x \forall y \, (Set(x) \wedge Set(y) \rightarrow (x = y \leftrightarrow \forall z \, (Sz \rightarrow (z \in x \leftrightarrow z \in y))))$$

i.e. that two sets are identical if they have the same members.

[32] Hale and Wright are not explicit whether the notion of an identity criterion is best understood as epistemological or metaphysical.

2. Arithmetic. It follows from the Peano axioms that

$$\forall x \forall y \, (\mathbb{N}x \wedge \mathbb{N}y \rightarrow (x = y \leftrightarrow \forall z \, (Pxz \leftrightarrow Pyz)))$$

   i.e. that two natural numbers are identical if they have the same predecessors.[33]
3. Real numbers. It follows from the axioms of suitable treatments of the real numbers that

$$\forall x \forall y \, (\mathbb{R}x \wedge \mathbb{R}y \rightarrow (x = y \leftrightarrow \forall z \, (\mathbb{Q}z \rightarrow (z < x \leftrightarrow z < y))))$$

   i.e. that two real numbers are identical if they form the same "cut" in the rational numbers.
4. Complex numbers. It follows from the axioms of suitable treatments of the complex numbers that

$$\forall x \forall y \, (\mathbb{C}x \wedge \mathbb{C}y \rightarrow (x = y \leftrightarrow (\Re(x) = \Re(y) \wedge \Im(x) = \Im(y))))$$

   where $\Re$ and $\Im$ are functions from $\mathbb{C} \rightarrow \mathbb{R}$ that respectively express the real and imaginary parts of a complex number.

More generally, introduce a new constraint on an acceptable theory intended to introduce some mathematical sort M—call it *Identity*—to the effect that the theory must entail a sentence of the form

$$\forall x \forall y \, (Mx \wedge My \rightarrow (x = y \leftrightarrow \Phi(x, y))$$

where Φ is a formula expressing an equivalence relation on M-objects.

The idea, in brief, is that theory-internal adjudications of identity are capable of playing the role generally assigned to identity criteria. More specifically, take a theory that entails a claim of this kind. This allows us to understand an instance of the Hilbertian Strategy as taking the axioms of the theory as an implicit definition *of a sortal concept*: the concept of the relevant kind of mathematical object (perhaps sets, natural or real numbers, etc). The way in which this handles identity claims should be clear. Identity claims between objects of the relevant sort are to be handled straightforwardly in terms of the particular theory-internal adjudication; whereas the issue of cross-sortal identity claims is dealt with in much the same way as on the orthodox neo-Fregean account (Hale and Wright (2001, 389): "across categories, objects are distinguished by just that—the fact that they belong to different categories.")

[33] An analogous condition could be written out, less perspicaciously, in terms of the successor function.

More needs to be said about what, exactly, is required in order for a theory to provide an internal adjudication of identity in the relevant sense. Ideally, it would be desirable to enumerate further (necessary and sufficient) conditions. Although I cannot offer anything like a full theory here, a number of plausible conditions can be made out, many of which can be adopted straightforwardly from the literature on criteria of identity.[34] For instance, Horsten (2010) mentions these:

1. Formal adequacy: a criterion of identity must express an equivalence relation.
2. Material adequacy: a criterion of identity must be true.
3. Necessity: a criterion of identity must follow from "theoretical principles concerning the subject matter in question".
4. Informativeness: a criterion of identity must be informative about the nature of the entities involved.
5. Non-circularity/Predicativity : a criterion of identity must not (essentially?) quantify over the entities whose identity is supposed to be established.

The last condition is somewhat tricky to formulate, if indeed it is a genuine constraint. However exactly it is done, it had presumably better not rule out the credentials of the Axiom of Extensionality, which is agreed on virtually all sides to be a paradigm case of an acceptable identity criterion.

At any rate, I do not propose that this list is exhaustive, and no doubt, more could be said. Nevertheless, I take it that it is extremely encouraging that all of the rough criteria set out above can be equally plausibly applied to theory-internal adjudications of identity. Whether or not one wishes to count them as criteria of identity proper, they each seem plausibly capable of performing the required job of introducing a genuine *sortal* concept, thereby allowing us a means of adjudicating identity claims between objects of the relevant sort and of explaining why identity claims between mathematical objects and objects of another sort—say, people—are unproblematically false.[35] If that is right, I can see no reason that they should not play roughly the role that identity-criteria play for the neo-Fregean. At the very least, we would need to hear a much more detailed case that the form of abstraction principles are uniquely suited to introducing sortal concepts than we have so far heard.

[34] I think that everything in the discussion above is consistent with the claim that what I have been calling theory-internal adjudications of identity simply are criteria of identity, though I would not like to make such a claim outright.

[35] The appeal to sortal concepts does not adjudicate the issue of identity claims between objects of putatively distinct *mathematical* sorts—e.g. the natural number 2 and the real number 2. This is a subtle issue, for neo-Fregeans as for others. See Fine (2002, I.5) for discussion. As far as I can tell, the proposal in Hale and Wright (2001, Ch 14) implies that identity claims of this kind are false. Thanks to a referee here.

5.4.1.2 Does the Hilbertian Strategy Attempt to Stipulate Truth?

It is worth briefly considering another complaint, related to the one discussed immediately above, where Hale and Wright appear to accuse the Hilbertian Strategist of attempting to *stipulate entities into existence*. Here is a representative passage:

> Before there is any question of anybody's knowing the vehicle to be true, the stipulation has first to make it true. Regrettably, we human beings are actually pretty limited in this department—in what we can make true simply by saying, and meaning: let it be so! No one can effectively make it true, just by stipulation, that there are exactly 200,473 zebras on the African continent. How is it easier to make it true, just by stipulation, that there is an ω-sequence of (abstract) objects of some so far otherwise unexplained kind? And even if we do somehow have such singular creationist powers, does anyone have even the slightest evidence for supposing that we do?... to lay down Dedekind-Peano as true is to stipulate, not truth-conditions, but truth itself.[36]

It seems to me that here, Hale and Wright are misled by their own rhetoric of "stipulation". What is going on, the Hilbertian theorist claims, is an instance of implicit definition, and it works no differently here than it does in the (putatively more favourable) case of Hume's Principle. A sentence containing previously uninterpreted vocabulary is being integrated into a pattern of usage in such a way as to give the sentence certain truth-conditions and the new vocabulary certain semantic values; no more, and no less. It is simply a mistake to suppose that anything objectionably "creationist" is under consideration.

Can anything more be said to support the accusation that Hilbertian implicit definitions involve the stipulation of *truth*, whereas abstraction principles merely assign *truth-conditions* to their left-hand-sides? On the neo-Fregean view, the success of HP (along with suitable definitions) ensures that, e.g. "0=0" has the same truth-conditions as the claim that the concept $x \neq x$ is in bijection with itself. Thus, on a coarse-grained view of "truth-conditions", HP ensures that this sentence has necessarily and always satisfied truth-conditions. If that is legitimate, however, there is no reason why the Hilbertian cannot claim the same status for the axioms of PA. Can the objection be pressed further if "truth-conditions" are understood in a more fine-grained way? The Hilbertian could, if necessary, take as the relevant implicit definition not the axioms of PA themselves but the biconditional consisting of the conjunction of the axioms on one side and some logical truth on the other. It might be objected in turn that this is inadequate as an implicit definition, since the choice of logical truth (and thus the truth-conditions assigned to the conjunction of the PA axioms) is arbitrary. Perhaps a natural candidate modification for the right-hand-side of the relevant implicit definition (inspired by the historical Hilbert himself) is the claim *that the axioms of* PA *are logically consistent*.[37]

[36] Hale and Wright (2009a, 473).

[37] Thanks to a referee for pressing for clarity here.

5.4.2 *Objections to the Epistemic Role of Hilbertian Definitions*

5.4.2.1 Is HP Objectionably Arrogant?

Let me now turn to a concern, raised by Hale and Wright, to the effect that the Hilbertian Strategy is arrogant. It proceeds by way of the subtle observation that the equivalence between the (theories resulting from the) Hilbertian Strategy and the neo-Fregean Strategy is contingent upon the choice of background logic. More precisely, the situation is this. If the underlying logic is classical second-order logic, then HP and PA will have precisely the same models.[38] However, consider what happens when we work within a weaker *Aristotelian* logic in which the second-order constants denote and second-order quantifiers range over only *instantiated* concepts. In such a setting no empty concept will be countenanced. This impedes Frege's theorem from going through at a very early stage: for in order to define the number zero as the cardinality of some empty concept, one must be available. So, as Wright and Hale note, taken against a background of Aristotelian logic, HP has models of both finite and infinite cardinality, while PA has—just as in the classical case—only infinite models. As they put it:

> The least one has to conclude from this disanalogy is that, as a stipulation, Hume is considerably more modest than Dedekind-Peano: the attempted stipulation of the truth of Dedekind-Peano is effectively a stipulation of countable infinity; whereas whether or not Hume carries that consequence is a function of the character of the logic in which it is embedded—and more specifically, a function of aspects of the logic which, one might suppose, are not themselves a matter of stipulation at all but depend on the correct metaphysics of properties and concepts.[39]

While the logical difference that Hale and Wright advert to is undeniable, the question is whether it can legitimately be used to support the claim that the Hilbertian Strategy is *objectionably arrogant*. This would require the demonstration of two subclaims: (i) that due to this logical difference, a stipulation of HP really is "more modest" than a stipulation of PA, and (ii) that this difference genuinely marks a difference in the acceptability of the two kinds of definition for the epistemic purposes to which they are intended to be put. It is worth emphasizing that (i) is not enough—it is clearly consistent to hold that a stipulation of HP is more modest than a stipulation of PA while maintaining that they both, so to speak, end up on the right side of the knowledge-conferring line.

To respond, I argue that abstractionists who seek to recover arithmetic cannot feasibly make this objection.[40] For as mentioned earlier, neo-Fregeans want to use HP plus second-order logic to derive the PA axioms and thereby put the whole of classical arithmetic within reach. And for just the reasons above, *orthodox* second-

[38] As usual, modulo appropriate definitions.

[39] Hale and Wright (2009a, 475).

[40] In keeping with the general methodology of the paper, I assume that something like the neo-Fregean account is tenable, putting aside more fundamental objections.

order logic is required here, for in an Aristotelian setting, that derivation will not go through. So, presumably, Hale and Wright must feel themselves entitled to appeal to orthodox second-order logic and thereby somehow to discount what would otherwise be the epistemic possibility that Aristotelian logic is ultimately the correct logic. If they are correct, then the logical difference between HP and PA to which they advert is simply irrelevant, because both neo-Fregeans and Hilbertians are justified in discounting that Aristotelian logic is appropriate for reasoning in the relevant setting. To put it another way: if it is a genuine epistemic possibility that the right logic is Aristotelian, then there may well be grounds for drawing a line between HP and PA; however, this epistemic possibility would itself undermine the prospect of grounding arithmetic in the former.

Let us consider another way of pressing the arrogance objection. Recall our sharpening of the notion above: an implicit definition S is arrogant if (i) there is some condition C that must be satisfied for the truth of S; (ii) we require justification that C is satisfied in order to have justification in S; and (iii) we have no justification that C is satisfied. Is there any case to be made that PA is arrogant but HP is not, with the claim that there exist infinitely many objects as the relevant condition C?

It is certainly true (assuming the legitimacy of orthodox second-order logic) that the existence of infinitely many objects is a condition for the truth of both HP and PA, so (i) holds for both approaches. Plausibly, too, proponents of both approaches are in a similar epistemic position (prior to the laying down of any implicit definition) regarding this condition, so that the situation with respect to its justification is also symmetric. The crux of the issue is whether there is any scope to argue for a difference in (ii), in particular, that justification in the existence of an infinity of objects is somehow required *in advance* of justifiably accepting PA, but not in advance of justifiably accepting HP. The only way I can see that this might be done is by arguing that the existence of infinitely many objects is a *immediate* consequence of PA, in some epistemically relevant sense of "immediate", whereas the same does not go for HP. But I cannot see any plausibility in the general principle that in order to have justification in P, one must have antecedent justification in its immediate consequences, on any precisification of the notion of immediacy. Furthermore, on this view, the arrogance objection would be easy to defeat by slightly modifying the statement of the Hilbertian Strategy. Take for instance the version of the view discussed in the last subsection, whereby the relevant implicit definition is not the axioms themselves but rather the biconditional consisting of the axioms on one side and the claim that the axioms are consistent on the other. It is very hard to see any sense in which this principle leads immediately to the consequence that there exist infinitely objects while Hume's Principle does not.

5.4.3 Objections Concerning the Applicability of Mathematics

A salient feature of the abstractionist strategy concerning at least arithmetic is that the applications of the theory come off-the-shelf. Hume's Principle, in effect, builds into the very identity conditions of numbers their role as, so to speak, measures of cardinality. By contrast, PA appears simply silent on the question of applicability; it says nothing, on its face, about how arithmetic can be applied. In particular, although PA delivers a body of truths concerning the natural numbers, it says nothing about how we are to decide what the number *of*, e.g., a concept might be.

With that said, it is not difficult to obtain a perfectly satisfactory "number of" operator, given the axioms of PA and the resources of implicit definition. An operator "#" can be defined with the truth-condition that

$$\#\Phi(x) = n \leftrightarrow \Phi(x) \approx \{x < n\}$$

where $\approx$ expresses equinumerosity in the usual second-order-logical way, $\{x < n\}$ expresses the property of being a number strictly less than n (where the less-than relation is defined in the language of PA as usual). Using standard second-order resources, this allows the derivation of Hume's Principle from PA. Consequently there should be no *technical* worries that the Hilbertian Strategy for arithmetic is somehow less able to provide for its applicability.

Nevertheless, a more fundamental objection is lurking. It might be argued that this opposition—placing the applicability-conditions of numbers at the center of the implicit definition by which they are introduced versus introducing them in some other way altogether—is *philosophically*, if not technically, significant. Many philosophers, Frege himself included, have placed the applicability of arithmetic at the heart of the subject. As Frege famously said, "…it is application alone that elevates arithmetic beyond a game to the rank of a science. So applicability necessarily belongs to it."[41]

For some clarity on the question, consider two salient families of positions that one might take concerning the applicability of mathematical theories. First, there are, following Pincock (2011, 282), what we might call *one-stage* views. On such a view, the applicability of mathematical theories is not merely a peripheral feature of them, something that arises as a happy coincidence once the theory has been formulated and worked out. Rather, it is *part of the content* of the theory that it can be applied in the way that it is. As Wright puts the idea—he calls it Frege's Constraint—"a satisfactory foundation for a mathematical theory must somehow build its applications, actual and potential, into its core—into the content it ascribes to the statements of the theory".[42]

[41] Frege et al. (2013, 100).

[42] Wright (2000, 324). For an attempt to provide a one-stage account of the real numbers, see Hale (2000). See also Batitsky (2002) for interesting arguments that Hale's view is inferior to

By contrast, consider *two-stage* views, according to which the application of mathematics can be understood as a two-step process. The first stage is an characterization of the subject-matter of (pure) mathematics that is autonomous of, i.e. makes no reference to, its applications. Examples of such views include straightforward platonism (according to which mathematics concerns a domain of mind-independent objects, picked out in a way that does not mention their applications), modal structuralism (Hellman 1989), *ante rem* structuralism (Shapiro 1997), and fictionalism (Field 2016).[43]

The second stage of a two-stage view then explains how, in light of the characterization of mathematics at the first stage, applications are possible. For instance, at a very high level of abstraction, this could be done by beginning with a purely mathematical domain and appealing to "isomorphism" or "representation" relations of structural similarity between the mathematical domain and the non-mathematical domain to which it is to be applied, thereby allowing the complex mathematical techniques and inference patterns that have been developed in pure mathematics to be brought to bear on problems concerning the structure of, e.g., the physical world.

The difference between the Hilbertian Strategy and the neo-Fregean Strategy for arithmetic can be seen as a microcosm of the opposition between one- and two-stage views. The neo-Fregean Strategy puts the possibility of counting objects at the core of the theory of arithmetic, in the strong sense that the very criterion of identity for numbers essentially involves their role as the measures of cardinality of concepts. By contrast, the Hilbertian Strategy is silent on the question of the application—counting—until it is augmented with the Dedekind-inspired definition of the 'number of' operator, which, in effect, introduces counting by setting up a "representation" relation between the objects falling under a certain concept and the natural numbers themselves.

In light of the foregoing distinction, the possibility of an argument against the Hilbertian Strategy and in favour of the neo-Fregean opens up—*if,* that is, Frege's Constraint holds. The same goes, *mutatis mutandis*, for other abstractionist treatments of mathematical theories: if a one-stage account of the applicability of the theory is plausibly required, and if the abstraction principle generating the theory can plausibly be said to place its applicability at the core of the account (in the way that HP does), then it seems the neo-Fregean Strategy will have demonstrable advantages over the Hilbertian Strategy. The task for the neo-Fregean is establishing the plausibility of these required premises. So, let us ask: is there any good *argument* for preferring a one-stage account over a two-stage account, in the case of arithmetic

---

the orthodox representation-theoretic explanation of the applicability of the reals (an explanation closer to the two-stage accounts I discuss below).

[43] Gianluigi Oliveri has helpfully pointed out that Brouwer's intuitionism is an additional example, since, for Brouwer, the activities of the mathematical Creative Subject have nothing to do with applicability (nor with language).

in particular and in other areas of mathematics in general?[44] The question is a large one, but in the remainder of this section I will consider some arguments that one-stage views are required and attempt to rebut them on behalf of two-stage views.

5.4.3.1 Can Only One-Stage Accounts Explain the Generality of Applications?

In a discussion of Frege, Dummett argues for something like a one-stage account:

> But the applicability of mathematics sets us a problem that we need to solve: what makes its applications possible, and how are they to be justified? We might seek to solve this problem piecemeal, in connection with each particular application in turn. Such an attempt will miss the mark, because what explains the applicability of arithmetic is a common pattern underlying all its applications. Because of its generality, the solution of the problem is therefore the proper task of arithmetic itself: it is this task that the formalist [i.e. the two-stage theorist], who regards each application as achieved by devising a new interpretation of the uninterpreted formal system *and as extrinsic to the manipulation system*, repudiates as no part of the duty of arithmetic[. . .]
>
> It is what is in common to all such uses, and only that, which must be incorporated into the characterisation of the real numbers as mathematical objects: that is how statements about them can be allotted a sense which explains their applications, without violating the generality of arithmetic by allusion to any specific type of empirical application.[45]

I take the argument in the foregoing to be this: only one-stage accounts of the applicability of various parts of mathematics (arithmetic and real analysis in particular) are able intelligibly to explain the possibility of the application of those theories in the required generality that the phenomenon requires.

In response: I want to say that it is glaringly unclear what is supposed to be lacking in rival, two-stage ("formalist") accounts. The charge is that a two-stage account can give, at best, a piecemeal explanation—missing, as Dummett puts it, the common pattern. Settle for the sake of argument on a platonist two-stage view of mathematics. (A similar account will be available, *mutatis mutandis*, for other two-stage views.) The explanation is, very roughly, that applications depend on structural iso- or homomorphisms between the mathematical objects and the non-mathematical objects that they purport to model. If successful, it explains not only why the specific non-mathematical domain in question can have mathematics applied to it, but equally well why any non-mathematical domain that is structurally similar can have mathematics applied in the same way also. What further common pattern is there to be explained?

Take the example of arithmetic once again as an illustration. Presumably, generality in Dummett's sense here is the datum that Frege himself emphasized

[44] For sophisticated recent discussions of the issues surrounding neo-Fregeanism and applications, congenial to the view proposed here, see Snyder et al. (2020), Panza and Sereni (2019), and Sereni (2019).

[45] Dummett (1991a, 259).

so heavily, namely that numbers can be assigned to any objects whatsoever (falling under a particular concept). Frege seeks to explain this with an account of numbers characterizing them in terms of their role in counting. But why think that such an account is the only sort able to explain the datum? Why can a two-stage account of counting not endorse and explain such a claim just as satisfactorily as the neo-Fregean one-stage account, as follows: all objects can be counted, simply because any plurality of objects can be linearly ordered and thereby, if finite, related isomorphically to some initial segment of the natural numbers. I do not see that this explanation lacks any required generality.

5.4.3.2 Are One-Stage Accounts Required to Do Justice to the *Practice* of Applications?

Dummett identifies a second, subtler Fregean argument for one-stage accounts of applicability. He writes:

> The historical genesis of the theory will furnish an indispensable clue to formulating that general principle governing all possible applications. ...Only by following this methodological precept can applications of the theory be prevented from assuming the guise of the miraculous; only so can philosophers of mathematics, and indeed students of the subject, apprehend the real content of the theory.[46]

This line of thought has been developed and expanded by Wright (2000).[47] Against roughly what I have called a two-stage view of applications, Wright offers an intriguing objection. He first notes that it is simply a datum that it is possible to come to appreciate simple truths of (e.g.) arithmetic via their applications. His primary example is a schoolyard demonstration that $4 + 3 = 7$ by simply counting on one's fingers. It is at least plausible that he is right here; surely this is at least one perfectly acceptable route to certain arithmetical knowledge, and indeed, the sort of route originally taken by many. But, the objection continues, it is difficult to square the existence of this route with a two-stage account of applications: it is not that the arithmetical knowledge is obtained by *first* apprehending the structure of the natural numbers and *then* drawing an inference that the fingers in question can be isomorphically related to some initial segment thereof; rather, it seems that by going through the counting routine one *thereby* grasps the arithmetical proposition itself, and this suggests that the content of the arithmetical proposition cannot be alienated from its application conditions in the way that the two-stage theorist contends.

For the sake of clarity, it's worth emphasizing the complaint is not that two-stage accounts misrepresent the actual order of understanding.[48] Rather, it's that

[46] Dummett (1991a, 300-1).

[47] Related issues are further discussed in Wright (2020).

[48] Pincock (2011) seems to read the objection in this way. But this cannot be what is meant, for even the neo-Fregean's own approach is *highly* intellectualized, and in particular, far too intellectualized to plausibly serve as a reconstruction of the actual order of understanding.

the two-stage account, even as an idealized, rational reconstruction of our practice, cannot allow that naive schoolyard demonstrations are demonstrations of genuinely *arithmetical* propositions. As Wright puts it, two-stage accounts of arithmetic "involve a representation of its content from which an appreciation of potential application will be an additional step, depending upon an awareness of certain structural affinities."[49] As a result, they seem open to the charge of changing the subject.[50]

I want to outline two possible lines of response here. First, and weakest, one might simply accept Wright's point and in response distinguish between different philosophical projects that one might engage in, differing over the centrality they accord to actual mathematical practice. We need to ask: what exactly do we want from a philosophical reconstruction of mathematics? There are different desiderata here, and they plausibly pull in different directions. On the one hand, one might seek an account of mathematics that is (at least approximately) faithful to the actual genealogy of mathematical belief and that, consequently, is able to explain why the schoolyard demonstration is a way of coming to know arithmetical propositions; and this might militate towards giving one-stage accounts of at least the most conceptually basic parts of mathematics. But that is not the only project one might be interested in. For on the other hand, one might instead emphasize the *uniformity* of mathematics, and seek to provide a homogeneous account that takes into account the most sophisticated modern conceptions of the various mathematical domains. This may well push us in the direction of giving two-stage accounts across the board, even at the expense of appearing to create a gulf between sophisticated mathematical beliefs and their more naive counterparts.

But there is a second line of response worth exploring, one that takes a less concessive tack and attempts to reconcile two-stage accounts with the datum that schoolyard demonstrations are (or at least can be) demonstrations of mathematical propositions. The crucial point to note here is that we seem to require a good deal of collateral conceptual mastery to count these basic counting routines as genuine demonstrations of arithmetical facts. What I mean by this is simply that we seem to require the performer of the routine to demonstrate an awareness that it is in a certain sense *general*, not merely to conclude on its basis that 4 *fingers* and 3 *fingers* are 7 *fingers*, but that 4 of *anything* plus 3 more make 7 things, *whatever things they may happen to be*. Getting genuine arithmetical knowledge from this kind of demonstration requires recognition that the particular features of the objects involved, aside from their cardinality, are irrelevant. (Consider how reluctant we would be to ascribe knowledge that $4 + 3 = 7$ to someone who goes through the routine but only seems to appreciate the identity as applied to the number of *fingers*.)

[49] Wright (2000, 327).

[50] I should emphasize that Wright does not argue that Frege's constraint holds across the board. Rather, he thinks, it holds when our initial understanding of a mathematical domain involves applications; but, in his view, when it does *not*—as is plausible for, e.g., complex analysis—then a two-stage account may well be acceptable.

I think this strongly suggests, *contra* Wright, that we *do* require an awareness of structural affinities—between e.g. the fingers, the numerals used, the natural numbers they refer to, and indeed any other objects of the same cardinality—in order for instances of the finger-counting routine to succeed in bestowing *arithmetical* knowledge (as opposed to merely knowledge about the particular fingers).[51] And if this is all right, then it is a non-sequitur to think that the two-stage explanation of applications cannot do justice to the phenomena.[52]

5.5 Conclusion

The main thrust of this essay has been that many of the subtle and ingenious resources devised by abstractionist neo-Fregeans to elaborate and defend their view can, with equal justice, be appropriated by neo-Hilbertians, who construe the axioms of mathematical theories themselves as implicit definitions. Furthermore, I've argued that such an approach has certain definite advantages over neo-Fregean abstractionism, and that it can plausibly respond to the main criticisms that have been raised against it. Perhaps these criticisms can be developed further, or additional ones can be pressed. But I hope that at the very least, the Hilbertian approach is seen as a plausible contender, worthy of further consideration for those who are tempted by the allure of neo-Fregeanism.[53]

Appendix

We prove a result, Lemma 1, which entails both Theorems 1 and 2.

[51] I am not saying that the *only* way of making sense of this generality requirement involves a recognition of structural affinities: indeed, neo-Fregeans offer an alternative explanation, since on their view the concept of number is given by HP, which itself provides the resources to attribute number to any concept. The dialectical point still stands, however: once it is recognized that schoolyard derivations involve further conceptual mastery, the immediate objection against two-stage accounts is substantially weakened. Thanks to a referee for discussion here.

[52] To avoid misunderstanding, my claim that this additional step—the recognition of structural affinities—is required is not a claim that the resulting belief that $4 + 3 = 7$ needs to be *inferred* from the schematic application to fingers. I am not making any claims about the architecture of inference at all. Rather, I am making claims about the architecture of *justification,* and it is this that I think supports, or at least is consistent with, the two-stage account.

[53] I am grateful to two anonymous referees for many constructive suggestions. Thanks to Andrea Christofidou, Hannes Leitgeb, Sabina Lovibond, Steven Methven, Beau Mount, Gianluigi Oliveri, Michail Peramatzis, Martin Pickup, Andrea Sereni, Stephen Williams, Crispin Wright, Luca Zanetti, and audiences at New York University, Worcester College, Oxford, the Munich Center for Mathematical Philosophy and IUSS-Pavia, for very helpful comments and discussion. Special thanks to Lavinia Picollo and Jared Warren for detailed comments on several drafts.

We work with second-order languages without function symbols and only monadic higher-order variables, for simplicity. Moreover, we assume the only logical symbols are $=, \neg, \vee$, and $\exists$. First-order languages are understood as fragments of a second-order language.

If $\mathscr{L}$ is a second-order language, model/interpretation $\mathcal{M}$ of $\mathscr{L}$ is just a model $\mathcal{M}$ of $\mathscr{L}$'s first-order fragment in which the higher-order quantifiers range over the power set of $\mathcal{M}$'s domain, $|\mathcal{M}|$.

Let $\mathscr{L}_a$ and $\mathscr{L}_b$ be second-order languages with no individual constants in common. Let P_a and P_b be monadic predicate symbols not occurring in $\mathscr{L}_a$ or $\mathscr{L}_b$. For each formula ϕ of $\mathscr{L}_a$, ϕ^* is the result of relativizing all quantifiers in ϕ to P_a, i.e.

$$\phi^* = \begin{cases} \phi & \text{if } \phi \text{ is atomic} \\ \neg\psi^* & \text{if } \phi := \neg\psi \\ \psi^* \vee \chi^* & \text{if } \phi := \psi \vee \chi \\ \exists v\,(P_a(v) \wedge \psi^*) & \text{if } \phi := \exists v\,\psi \\ \exists V\,(\forall v\,(Vv \rightarrow P_a(v)) \wedge \psi^*) & \text{if } \phi := \exists V\,\psi \end{cases}$$

In the last clause, v is an individual variable not occurring in ϕ. Analogously, if ϕ is a formula of $\mathscr{L}_b$, ϕ^* is the result of relativizing all quantifiers in ϕ to P_b. Let $\mathscr{L}^*$ extend $\mathscr{L}_a \cup \mathscr{L}_b$ with P_a and P_b. If Γ_a is a set of sentences of $\mathscr{L}_a$ and Γ_b a set of sentences of $\mathscr{L}_b$, then Γ_a^* is the set of sentences ϕ^* of $\mathscr{L}^*$ such that $\phi \in \Gamma_a$, and similarly for Γ_b^*.

Lemma 1 *If Γ_b is satisfiable and, for some $\phi \in \mathscr{L}_a$ ($\phi \in \mathscr{L}_b$), $\Gamma_a^* \cup \Gamma_b^* \vDash \phi^*$, then $\Gamma_a \vDash \phi$ ($\Gamma_b \vDash \phi$).*

Proof Assume Γ_b is satisfiable, so it must have a model, $\mathcal{M}_b$. Assume, for reductio, that for some sentence $\pi \in \mathscr{L}_a$, $\Gamma_a^* \cup \Gamma_b^* \vDash \pi^*$ but $\Gamma_a \nvDash \pi$. Thus, $\Gamma_a \cup \{\neg\pi\}$ must have a model $\mathcal{M}_a$ which is, trivially, a model of Γ_a too. We establish the result by showing that $\mathcal{M}_a$ and $\mathcal{M}_b$ can be extended to a model $\mathcal{M}^*$ of $\Gamma_a^* \cup \Gamma_b^*$; thus we have $\Gamma_a^* \cup \Gamma_b^* \nvDash \pi^*$, contradicting our assumption.

Let $\mathcal{M}^*$ be the following model of $\mathscr{L}^*$:

1. $|\mathcal{M}^*| = |\mathcal{M}_a| \cup |\mathcal{M}_b|$.
2. If c is an individual constant of $\mathscr{L}_a$, then $c^{\mathcal{M}^*} = c^{\mathcal{M}_a}$, and similarly for $\mathscr{L}_b$. This is always possible because $\mathscr{L}_a$ ad $\mathscr{L}_b$ don't share any individual constants.
3. If P is a relation symbol exclusively of $\mathscr{L}_a$, $P^{\mathcal{M}^*} = P^{\mathcal{M}_a}$, and similarly for $\mathscr{L}_b$. If P occurs both in $\mathscr{L}_a$ and in $\mathscr{L}_b$, $P^{\mathcal{M}^*} = P^{\mathcal{M}_a} \cup P^{\mathcal{M}_b}$.
4. $P_a^{\mathcal{M}^*} = |\mathcal{M}_a|$ and $P_b^{\mathcal{M}^*} = |\mathcal{M}_b|$.

If σ is an assignment over a model $\mathcal{M}$ and $d \in |\mathcal{M}|$, σ_v^d is identical to σ except it maps the individual variable v to d. Similarly, if $S \in \mathscr{P}(|\mathcal{M}|)$, σ_V^S is identical to σ except it maps the second-order variable V to S.

Note that, by 1, every assignment σ over $\mathcal{M}_a$ or $\mathcal{M}_b$ is an assignment over $\mathcal{M}^*$. We prove without loss of generality, by induction on the logical complexity of the formula $\phi \in \mathscr{L}_a$, that, for every assignment σ over $\mathcal{M}_a$, $\mathcal{M}_a, \sigma \vDash \phi$ if and only if $\mathcal{M}^*, \sigma \vDash \phi^*$.

Note that this would suffice to establish our lemma. For every $\phi \in \Gamma_a$, $\mathcal{M}_a \vDash \phi$, the induction will yield that, for every $\phi^* \in \Gamma_a^*$, $\mathcal{M}^* \vDash \phi^*$; an analogous result for Γ_b can be proved in a similar fashion, so $\mathcal{M}^* \vDash \Gamma_a^* \cup \Gamma_b^*$. Moreover, since $\mathcal{M}_a \vDash \neg\pi$, we must have that $\mathcal{M}^* \vDash \neg\pi^*$, which means that $\Gamma_a^* \cup \Gamma_b^* \nvDash \pi^*$.

- If ϕ is an atomic formula, then $\phi^* = \phi$. Since σ assigns objects taken from $|\mathcal{M}_a|$ to each variable of $\mathscr{L}^*$, by 2 and 3, $\mathcal{M}_a, \sigma \vDash \phi$ iff $\mathcal{M}^*, \sigma \vDash \phi^*$.
- Assume the claim holds of every formula of lower complexity than ϕ.
  - If $\phi := \neg\psi$, then $\phi^* = \neg\psi^*$. Thus, $\mathcal{M}_a, \sigma \vDash \phi$ iff $\mathcal{M}_a, \sigma \nvDash \psi$ iff, by inductive hypothesis, $\mathcal{M}^*, \sigma \nvDash \psi^*$ iff $\mathcal{M}^*, \sigma \vDash \phi^*$.
  - If $\phi := \psi \vee \chi$, then $\phi^* = \psi^* \vee \chi^*$. Thus, $\mathcal{M}_a, \sigma \vDash \phi$ iff $\mathcal{M}_a, \sigma \vDash \psi$ or $\mathcal{M}_a, \sigma \vDash \chi^*$ iff, by inductive hypothesis, $\mathcal{M}^*, \sigma \vDash \psi^*$ or $\mathcal{M}^*, \sigma \vDash \chi^*$ iff $\mathcal{M}^*, \sigma \vDash \phi^*$.
  - If $\phi := \exists x\, \psi$, then $\phi^* = \exists x\, (P_a(x) \wedge \psi^*)$. Thus, $\mathcal{M}_a, \sigma \vDash \phi$ iff there is an $d \in |\mathcal{M}_a|$ s.t. $\mathcal{M}_a, \sigma_x^d \vDash \psi$ iff, by inductive hypothesis, $\mathcal{M}^*, \sigma_x^d \vDash \psi^*$ iff, by 4, $\mathcal{M}^*, \sigma \vDash \phi$.
  - Let $\phi := \exists X\, \psi$, then $\phi^* = \exists X(\forall x(Xx \rightarrow P_a(x)) \wedge \psi^*)$. Thus, $\mathcal{M}_a, \sigma \vDash \phi$ iff there is an $S \subseteq |\mathcal{M}_a|$ s.t. $\mathcal{M}_a, \sigma_X^S \vDash \psi$ iff, by inductive hypothesis, $\mathcal{M}^*, \sigma_X^S \vDash \psi^*$ iff, by 4, $\mathcal{M}^*, \sigma \vDash \phi$.

□

Note that Lemma 1 directly entails Theorem 2. Moreover, Theorem 1 also follows. For if the theories in question are first-order, then by the completeness theorem, (a) both theories are satisfiable iff they are consistent, and (b) one theory is a Field*-conservative extension of the other iff it is a Field-conservative extension; note that the last case of the previous induction is not relevant for such theories.

References

Batitsky, V. 2002. Some measurement-theoretic concerns about Hale's 'reals by abstraction'. *Philosophia Mathematica* 10(3): 286–303.

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

Burgess, J. 2004. E pluribus unum: Plural logic and set theory. *Philosophia Mathematica* 12(3): 193–221.

Burgess, J. 2005. *Fixing Frege*. Princeton: Princeton University Press.

Button, T., and S. Walsh. 2018. *Philosophy and Model Theory*. Oxford University Press.

Cook, R. 2016. Conservativeness, cardinality, and bad company. In *Abstractionism*, ed. P. Ebert and M. Rossberg, 223–246. Oxford University Press.

Dummett, M. 1991a. *Frege: Philosophy of Mathematics*. Cambridge: Harvard University Press.

Dummett, M. 1991b. *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.

Ebert, P., and S. Shapiro. 2009. The good, the bad and the ugly. *Synthese* 170(3): 415–441.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
Field, H. 1984. Platonism for cheap? Crispin Wright on Frege's context principle. *Canadian Journal of Philosophy* 14: 637–662.
Field, H. 2016. *Science Without Numbers*, 2nd ed. Oxford: Oxford University Press.
Fine, K. 2002. *The Limits of Abstraction*. Oxford University Press.
Frege, G. 1982. *Philosophical and Mathematical Correspondence*. Oxford: Blackwell.
Frege, G., P.A. Ebert, and R.T. Cook. 2013. *Gottlob Frege: Basic Laws of Arithmetic*. Oxford: Oxford University Press.
Hale, B. 2000. Reals by abstraction. *Philosophia Mathematica* 8(2): 100–123.
Hale, B., and C. Wright. 2000. Implicit definition and the a priori. In *New Essays on the A Priori*, ed. P. Boghossian and C. Peacocke, 286–319. Oxford: Oxford University Press.
Hale, B., and C. Wright. 2001. *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Oxford University Press.
Hale, B., and C. Wright. 2009a. Focus Restored: Comments on John MacFarlane. *Synthese* 170(3): 457–482.
Hale, B., and C. Wright. 2009b. The metaontology of abstraction. In *Metametaphysics: New Essays on the Foundations of Ontology*, ed. D. Chalmers, D. Manley, and R. Wasserman, 178–212. Oxford: Oxford University Press.
Heck, R.K. 2000. Cardinality, counting, and equinumerosity. *Notre Dame Journal of Formal Logic* 41(3): 187–209.
Hellman, G. 1989. *Mathematics Without Numbers: Towards a Modal-Structural Interpretation*. Clarendon Press.
Horsten, L. 2010. Impredicative identity criteria. *Philosophy and Phenomenological Research* 80(2): 411–439.
Linnebo, Ø. 2009. Introduction [special issue on the Bad Company objection]. *Synthese* 170: 321–329.
MacBride, F. 2003. Speaking with shadows: A study of neo-logicism. *The British Journal for the Philosophy of Science* 54(1): 103–163.
MacFarlane, J. 2009. Double vision: Two questions about the neo-fregean program. *Synthese* 170(3): 443–456.
Panza, M., and A. Sereni. 2019. Frege's constraint and the nature of Frege's foundational program. *Review of Symbolic Logic* 12(1): 97–143.
Pincock, C. 2011. *Mathematics and Scientific Representation*. Oxford: Oxford University Press.
Sereni, A. 2019. On the philosophical significance of Frege's constraint. *Philosophia Mathematica* 27(2): 244–275.
Shapiro, S. 1991. *Foundations Without Foundationalism: A Case for Second-Order Logic*. Oxford: Oxford University Press.
Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.
Shapiro, S., and A. Weir. 1999. New V, ZF and abstraction. *Philosophia Mathematica* 7(3): 293–321.
Snyder, E., R. Samuels, and S. Shapiro. 2020. Neologicism, Frege's constraint, and the Frege-Heck condition. *Noûs* 54(1): 54–77.
Tennant, N. 1978. *Natural Logic*. Edinburgh: Edinburgh University Press.
Wright, C. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
Wright, C. 1999. Is Hume's principle analytic? *Notre Dame Journal of Formal Logic* 40(1): 6–30.
Wright, C. 2000. Neo-Fregean foundations for real analysis: Some reflections on Frege's constraint. *Notre Dame Journal of Formal Logic* 41(4): 317–334.
Wright, C. 2007. On quantifying into predicate position: Steps towards a new(tralist) perspective. In *Mathematical Knowledge*, ed. M. Leng, A. Paseau, and M. Potter, 150–174. Oxford University Press.

Wright, C. 2016. Abstraction and epistemic entitlement: On the epistemological status of Hume's principle. In *Abstractionism*, ed. P. Ebert and M. Rossberg, 161–185. Oxford: Oxford University Press.

Wright, C. 2020. Is There Basic A Priori Knowledge of Necessary Truth? Elementary Arithmetic as a Case Study.

Part II
Structures and Structuralisms

Chapter 6
Structural Relativity and Informal Rigour

Neil Barton

Abstract Informal rigour is the process by which we come to understand particular mathematical structures and then manifest this rigour through axiomatisations. Structural relativity is the idea that the kinds of structures we isolate are dependent upon the logic we employ. We bring together these ideas by considering the level of informal rigour exhibited by our set-theoretic discourse, and argue that different foundational programmes should countenance different underlying logics (intermediate between first- and second-order) for formulating set theory. By bringing considerations of perturbations in modal space to bear on the debate, we will suggest that a promising option for representing current set-theoretic thought is given by formulating set theory using quasi-weak second-order logic. These observations indicate that the usual division of structures into *particular* (e.g. the natural number structure) and *general* (e.g. the group structure) is perhaps too coarse grained; we should also make a distinction between *intentionally* and *unintentionally* general structures.

Keywords Set theory · Continuum hypothesis · Higher-order logic · Informal rigour

6.1 Introduction

Mathematicians are often concerned with elucidating structure. In this paper, I'll examine some issues arising under the following assumption:

> (**Weak Structuralist Assumption**) Part of mathematics and its practice can be understood as isolating and studying different structures.

N. Barton (✉)
Fachbereich Philosophie, University of Konstanz, Konstanz, Germany
e-mail: neil.barton@uni-konstanz.de

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_6

Why is this assumption 'weak'? Well, the usual statement of structuralism is that mathematics just *is* the study of structure.[1] We do not make such a strong claim. Rather, we are just assuming the highly plausible claim that mathematics is at least partly concerned with the specification and study of structure.

Two questions are immediately pertinent:

1. What kinds of different structures are there?
2. How to we isolate them and/or talk about them?

The first question is often answered by distinguishing between two kinds of structure; *particular* and *general*. Isaacson explains the distinction as follows:

> The particularity of a particular structure consists in the fact that all its exemplars are isomorphic to each other. The generality of a general structure consists in the fact that its various exemplars need not be, and in general are not, isomorphic to each other. (Isaacson, 2011), p. 21

Exactly what different branches of mathematics have an underlying 'particular structure' is a contentious issue (we discuss this later). However, almost everyone agrees that we can talk about various kinds of *finite* particular structure (e.g. the structure of ten objects under some well-order). Normally it is assumed that most of our arithmetical talk is concerned with a particular structure; the standard model of arithmetic.[2]

General structures, by contrast, are not determined up to isomorphism and include groups, rings, and fields. An example: The group of symmetries on a triangle and the group of integers—both possess the general structure of being a group, but the former is finite where the latter is infinite.

It is a somewhat controversial question as to whether these two kinds of structure are of the same ontological kind or not, since particular structures *seem* more fundamental than general structures in the sense that the latter are properties that the former can possess. We speak, for example, of the particular structure of the integers *exemplifying* the ring structure or the particular structure of the natural

[1] A good example here is Shapiro:

> For our first (or second) approximation, then, pure mathematics is the study of structures, independently of whether they are exemplified in the physical realm, or in any realm for that matter. (Shapiro, 1997, p. 75)

Examples can be multiplied (e.g. Resnik 1997 and Hellman 1996). More generally, structuralist ideas have a rich history, appearing in the axiomatic work of Hilbert, Dedekind, and (under one interpretation) Zermelo. A different direction to the mathematical appeal of Structuralism was through the study of abstract algebra and related fields in the work of (among others) Bourbaki, Ore, and Noether (as well as contributions by Hilbert and Dedekind in this field as well), before the emergence of category theory and contemporary structuralist programmes in philosophy. See Corry (2004) for an in depth study of the history, and Reck and Schiemer (2019) for a survey of the state of the art.

[2] See Hamkins (2012) for a dissenting voice that we discuss a bit later.

numbers under addition *exemplifying* the general structure of a monoid.[3] Still more concrete are the *systems* exemplifying particular structures. For example, the face of the clock on my wall (with the usual operations of addition) is a *system* exemplifying the *particular* structure of the integers mod 12, which in turn exemplifies the *general* group structure.

The second question (how we isolate and talk about the different kinds of structure) is then easy in the case of general structures for the Weak Structuralist; she can simply state the conditions she is interested in for some general structure, and in doing so talks about any particular structures and/or systems that satisfy these conditions. The question is harder for particular structures, since here there is the additional challenge of convincing ourselves that we have isolated a structurally unique entity (at least up to isomorphism[4]). If a discipline or syntactic theory has a unique particular structure underlying it, then it is often referred to as a *non-algebraic* theory or discipline, those with no corresponding particular structure (or a general structure) are called *algebraic*.[5]

One way of tackling the question of when we have isolated a particular structure can be derived from the work of Kreisel (1967) and has been taken up subsequently by Isaacson (2011). They suggest that we have a process of *informal rigour* by which we obtain mathematical understanding and isolate different particular structures. The rough idea (which I discuss in more detail below) is that we isolate a particular structure by becoming more rigorous about a topic, and manifest this rigour by providing a categorical axiomatisation.

A categorical axiomatisation is a set of axioms **T** which determine a unique model up to isomorphism (i.e. any two models of **T** are isomorphic). Where categoricity is concerned, one must talk about different logics. The insight provided by the Löwenheim-Skolem Theorems shows that first-order logic cannot provide

[3] Isaacson (2011) seems to take the view that particular structures are somehow more fundamental, referring to a general structure with no particular instances as "vacuous" (p. 25). Similar remarks can be found in Leitgeb (2020), where unlabelled graphs are taken as the particular 'ground level' structures, and general structures are viewed as higher-order properties or classes of particular structures.

[4] There is a substantial discussion around whether isomorphism is too strong, and perhaps something weaker like *definitional equivalence* would be better. We set aside this issue for now, things are complicated enough without opening that can of worms, despite its interest. For an overview, see Button and Walsh (2018), Ch. 5.

[5] We discuss these distinctions in Sect. 6.2 below. The algebraic vs. non-algebraic distinction goes back at least to Shapiro (1997, pp. 40–41). Geoffrey Hellman points out that one might wish to eschew the use of the terms 'particular' and 'general' when discussing structures in favour of only talking about algebraic and non-algebraic theories. For the purposes of this paper, I will talk about both particular/general structures and algebraic/non-algebraic theories, since (a) some authors (e.g. Isaacson 2011) do use this terminology, (b) locutions like "the integers mod 12 exemplify the group structure" do not seem *obviously* impermissible, and (c) nothing too much hangs on this distinction for the purposes of the paper: My main aim is to analyse how our thought and language interacts with truth values for different claims—the theorist who wishes to eliminate talk of different kinds of structure is welcome to re-read the paper attending only to claims about truth values rather than the taxonomy of structures.

categorical axiomatisations for infinite structures. It is in the work of Resnik (in particular Resnik 1997) where we find a notion of *structural relativity*; the idea that the structure isolated for different parts of mathematics depends on the logical resources we consider.

This paper brings together these ideas focussing on set theory as a case study. We argue for the following claims:

1. For different foundational programmes corresponding to different levels of informal rigour, it is reasonable to hold that our set-theoretic thought is underwritten by in a logic *stronger* than first-order, but *weaker* than second-order.
2. This shows that the usual distinction between *particular* and *general* structures corresponding to different concepts is more fine-grained than we might have initially thought. There are concepts that correspond to *intentionally general structures* in that the concept is designed to talk about many non-isomorphic structures. Other concepts correspond to *unintentionally* general structures, where we do not *intend* for the structure we talk about to be general, yet we do not pin down a *particular* structure with our discourse.
3. We have some reason to doubt that we are *fully* informally rigorous about set theory. Rather, we might hold that our level of informal rigour is *partial*, and in particular our level of informal rigour is not yet enough to determinate a truth-value for the Continuum Hypothesis (CH).

Here's the plan: Sect. 6.2 examines the notion of informal rigour as it appears in Kreisel's (1967) paper and how it relates to the problem of the Continuum Hypothesis. We'll make the distinction between *top-down* informal rigour concerning particular structures (axiomatisations serve as certifications that informal rigour has been achieved) and *bottom-up* informal rigour (*given* an axiomatisation we use it to characterise particular structures). Section 6.3 presents three possible interpretations of informal rigour; a quasi-idealist one, a weakly platonistic one, and a strongly platonistic one (we'll see shortly what I mean by these terms). Section 6.4 presents the idea of structural relativity. Section 6.5 then examines different states we may be in with respect to informal rigour on the basis of different foundational programmes, and examines some possibilities for axiomatisations of our thought. We develop an assumption of Modal Definiteness; that informal rigour about a certain subject matter should not permit conceptual refinement motivating radically different axiomatisations (given a perturbation in temporal or modal space) and use this to analyse our level of informal rigour. Section 6.6 examines some objections and replies. In responding to possible objections, I develop a quadrilemma for the believer that CH has a determinate truth value; either (i) we mystically do not go astray when coming to justify new axioms, or (ii) we accept that we cannot justify new set-theoretic axioms, or (iii) it is possible to become *less* precise about the structure we talk about as we come to accept more axioms, or (iv) we have to give up a principle of charity in interpreting set-theoretic claims. Finally Sect. 6.7 concludes with some open questions.

6.2 Informal Rigour and the Continuum Hypothesis

In this section we explain *informal rigour* and the idea that it might be used to show the existence of particular structures. We'll do this by explaining Kreisel's rough idea, and then formulating a more precise thesis (that particular structures are determined via informal rigour) at the end of the section. We'll also explain how Kreisel thought that his account of informal rigour leads to a determinate truth value for the Continuum Hypothesis.[6]

Kreisel (1967) discusses the notion of *informal rigour*. This represents a development and refinement of the idea that we work mathematically by examining our intuitive notions and laying down axioms for them. Kreisel expands this thought by arguing that the process is not quite so simple; rather than merely analysing our intuitive concepts, we can become successively clearer about a mathematical subject matter and then manifest this clarity through axiomatisations. He writes:

> Informal rigour wants (i) to make this analysis [of intuitive notions] as precise as possible (with the means available), in particular to eliminate doubtful properties of the intuitive notions when drawing conclusions about them; and (ii) to extend this analysis, in particular not to leave undecided questions which can be decided by full use of evident properties of these intuitive notions. (Kreisel, 1967, pp. 138–139)

Kreisel's point is well-taken, and the history of mathematics is replete with notions that were initially unclear but slowly came to be made precise through development and reflection. Examples include ideas of completeness/continuity and denseness (early on these were sometimes confused), the notion of derivative (we will discuss this later in Sect. 6.6), Cantor's analysis of the size of sets, and indeed the notion of set itself was gradually made clearer. However, whilst Kreisel's remarks are suggestive, he does not provide a detailed account of exactly what informal rigour is like. Largely speaking, he takes it for granted that we know what it is when we see it (at least as far as his Kreisel (1967) is concerned).

Despite this, we can make some progress by examining specific questions:

(1.) What are the targets of informal rigour?[7]
(2.) How do we achieve informal rigour?
(3.) What are the consequences of informal rigour?

---

[6] Interestingly, it certainly seems like Kreisel held something like the Weak Structuralism. For example, he writes:

> if one thinks of the axioms as *conditions* on mathematical objects, i.e. on the structures which satisfy the axioms considered, these axioms make a selection *among* the basic objects; they do not tell us what the basic objects are. (Kreisel, 1967, p. 165, emphasis original)

Whilst the extent to which Kreisel *really was* a structuralist (rather than merely provided resources *useful* to structuralism) is certainly an interesting question, I lack the space to address it fully here.

[7] I thank Verena Wagner for pressing this question in discussion.

For (1.) some taxonomy will be useful. When we talk about mathematical structure, there are several important aspects:

(a) The *concepts* we employ in thinking about mathematics (I'll refer to these using C, C_0, C_1, ... etc.).[8]
(b) The *mathematised natural language(s)* we use when speaking about structure(s). We will refer to these as *discourses*, and denote them by (D, D_0, D_1, ...).
(c) Different formal mathematical *theories* ($\mathbf{T}$, $\mathbf{T}_0$, $\mathbf{T}_1$, ...).
(d) Different mathematical *structures*, both particular and general (S, S_0, S_1, ...).
(e) Different *systems* exemplifying structures, which for convenience we'll assume are model-theoretic structures ($\mathfrak{M}$, $\mathfrak{M}_0$, $\mathfrak{M}_1$, ...).

It is important to be clear about these distinctions if we are to provide a fully worked-out account on Kreisel's behalf. Nowhere is he fully explicit about the matter, but his discussion (and a reasonable understanding of the notion) seems to suggest that informal rigour concerns how the concepts underlying discourses can be refined in coming to be precise about structures. Mathematical practice involves communicating in a mathematised natural language, and how we interpret this language is contingent upon the concepts being employed. For example, the interpretation we ascribe to a computer scientist using the term "set" (in a context where we can have non-well-founded 'sets') is different from the interpretation we would ascribe to a set theorist working in some extension of **ZFC**. This isn't a contradiction; they are simply employing different concepts with their use of language and *mean* different things with their usage of the term "set". Correspondingly, there are different ways we could systematise or represent their language formally, and in turn different interpretations of this formal language. At the bottom level, the formal theories representing different pieces of mathematised language can be interpreted (contingent on the concepts employed) as about different kinds of structure.

In the rest of the paper, we will assume that the main *target* of informal rigour is the *concepts* we employ when speaking or writing in mathematical language (i.e. *discourses*). Perhaps there is more to be said here, but I'm happy to make this assumption for the purposes of the paper.

With the targets and rough idea of informal rigour in play, we can begin to address (2.) How do we achieve informal rigour? Kreisel provides four examples,[9]

[8] Juliette Kennedy suggests that talk of concepts is too unclear, and we would be better off eliminating this language altogether. I am somewhat sympathetic to this position, and certainly feel that it can sometimes muddy the waters. Despite this, language of this kind is useful for setting up the debate, and so I'll continue to use it here. For the reader who has doubt about the coherence of concept-talk, I suggest that they read all mention of concepts as shorthand for their favourite account of the constituents of thoughts.

[9] These include: (I) analysing the difference between independence results, such as the parallels axiom in geometry and the independence of CH in set theory (the focus of this paper), (II) the relation between intuitive consequence and syntactic/semantic consequence (here he gives his famous 'squeezing' argument, arguing that the informal notion of consequence can be squeezed

key to each is the idea that we develop informal rigour concerning a concept via working with it in practice. In this way we can develop our intuitions, and come to be rigorous about a notion. This rigour can then be formally codified. Our interest will be especially in his remarks about the difference between the independence of the Parallels Postulate from the second-order axioms of geometry, and the independence of CH from the axioms of **ZFC**.

Concerning the axioms of $\mathbf{ZFC}_2$,[10] Kreisel discusses the following:

Theorem (Zermelo 1930) [11] *Let* $\mathfrak{M}$ *and* $\mathfrak{N}$ *be models of* $\mathbf{ZFC}_2$. *Then either:*

1. $\mathfrak{M}$ *and* $\mathfrak{N}$ *are isomorphic.*
2. $\mathfrak{M}$ *is isomorphic to proper initial segment of* $\mathfrak{N}$*, of the form* V_κ *for inaccessible* κ.
3. $\mathfrak{N}$ *is isomorphic to proper initial segment of* $\mathfrak{M}$*, of the form* V_κ *for inaccessible* κ.

The core point is the following; whilst there is no full categoricity theorem for second-order set theory $\mathbf{ZFC}_2$, there is for initial segments.[12] In particular, many versions of $\mathbf{ZFC}_2$ with a specific bound on the number of large cardinals (e.g. "There are no inaccessible cardinals" or "There are exactly five inaccessible cardinals") *are* categorical.

Concerning this theorem, Kreisel writes:

> the actual formulation of axioms played an auxiliary rather than basic role in Zermelo's work: the intuitive analysis of the crude mixture of notions, namely the description of the type structure, led to the good axioms: these constitute a record, not the instruments of clarification. (Kreisel, 1967, p. 145)

How might we then determine a particular structure according to Kreisel? Abstractly speaking, Kreisel's position might then be described as follows. We begin to work with an informal concept C, employing it in some mathematical discourse D. Gradually we begin to become clearer about D and C via using them in practice, and developing our intuitions about the subject. Once we are eventually clear about

---

between the formal classes of a syntactic derivation in first-order logic and semantic consequence in first-order logic), (III) Brouwer's 'empirical' propositions, and (IV) showing that the use of certain models is a conservative extension of arithmetic.

[10] $\mathbf{ZFC}_2$ denotes the second-order formulation of **ZFC**, where the Axiom Scheme of Replacement is replaced with a single axiom quantifying over functions, and where the Axiom of Choice is replaced with the second-order claim that the universe can be well-ordered by a class-sized function. A concise presentation, including the quasi-categoricity result I discuss, is provided in Hekking (2015).

[11] Shepherdson (1951, 1952, 1953) takes Zermelo (1930)'s proof and clears up a few details. A modern presentations of proofs are available in Hekking (2015) and Button and Walsh (2018, §8A), and a version of Zermelo's proof in modern notation is available in Kanamori (2004). A different method, developed recently by Väänänen and Wang (2015) is to move to a proof-theoretic characterisation of categoricity (so called *internal* categoricity). We will discuss this move later in Sect. 6.6. See Button and Walsh (2018, Ch. 11) for an overview of the internal categoricity results.

[12] One can obtain a full categoricity proof of sorts with further meta-theoretic assumptions. See McGee (1997) for a full categoricity result using urelements. Since the assumptions required for this result are relatively controversial (see e.g. Rumfitt 2015, pp. 273–275) we set it aside here.

the right concept C' underlying D (it is at least possible that $C' = C$ here), we will have obtained sufficient precision to lay down a theory $\mathbf{T}$ for C', which is *categorical* in that any system $\mathfrak{M} \models \mathbf{T}$ is isomorphic to any other system $\mathfrak{M}' \models \mathbf{T}$. In this way, by employing our concept C' and using $\mathbf{T}$, we have determined a particular structure S up to isomorphism. In the case of set theory, we can think of the development of the idea of *cumulative hierarchy* and *iterative conception of set* after 1900 as yielding some particular set-theoretic structures by 1930 when Zermelo showed his 1908 axiomatisation was categorical. We will refer to the way that we can successively become clear about a concept determining a particular structure, before manifesting this rigour via a categorical axiomatisation as *top-down* informal rigour.

Top-down informal rigour is a way of coming to be clear about a concept and extracting an axiomatisation that determines a particular structure up to isomorphism. However it is not the only way that we can determine particular structures. Once we have accepted some logical resources and mathematical theory (possibly on top-down grounds) we can use these resources to determine other particular structures. For example, suppose that we have accepted informal rigour concerning the concept *natural number* and that $\mathbf{PA}_2$ manifests this informal rigour concerning a single unique structure (via the Dedekind categoricity theorem). We can then be informally rigorous about the concept *hereditarily finite set*, since we can find an interpretation of the hereditarily finite sets in the standard model of $\mathbf{PA}_2$, and this interpretation determines the hereditarily finite sets up to isomorphism.[13] But we *needn't* have been *already* informally rigorous about the concept *hereditarily finite set*, we have used other resources to characterise it. We will refer to informal rigour obtained via other accepted resources as *bottom-up* informal rigour concerning a concept.

This brings us on to (3.) What are the consequences of informal rigour? Our focus will be how informal rigour affects our attitude to the truth value of CH. Key here is the Zermelo quasi-categoricity theorem; this shows that *given an interpretation of the second-order variables* (this will be important later), $\mathbf{ZFC}_2$ determines several particular structures corresponding to initial segments of the cumulative hierarchy.

Kreisel took this to show that our talk concerning the cumulative hierarchy, as axiomatised by $\mathbf{ZFC}_2$, was unambiguous. He writes:

> Denying the (alleged) bifurcation or multifurcation of our notion of set of the cumulative hierarchy is nothing else but asserting the properties of our intuitive conception of the cumulative type-structure mentioned above. (Kreisel, 1967, pp. 144–145)

Why is this significant for CH? Well, since the truth value of CH is settled by $V_{\omega+2}$ (well below the least inaccessible) and if we think that all models of $\mathbf{ZFC}_2$ agree up to the first inaccessible (by the Zermelo quasi-categoricity theorem), then CH has the same truth-value in all particular structures meeting our informally

[13] The interpretation is via the Ackermann encoding of $\langle HF, \in \rangle$ into arithmetic. Sets are represented by natural numbers, and nEm when the nth binary digit of m is 1. $\langle \mathbb{N}, E \rangle$ is then isomorphic to $\langle HF, \in \rangle$. Of course, we can then give a categorical axiomatisation of the hereditarily finite sets using the theory $\mathbf{ZFC}_2$-Infinity+"There are no infinite sets".

rigorous concept of set (so the thinking goes). This, as Kreisel points out, makes the independence of CH from set-theoretic axioms markedly different from the independence of the Parallels Postulate (PP) from the axioms of geometry; PP can have different truth-values across models of the *second-order* axioms of geometry (once we fix upon some interpretation of the second-order variables), whereas CH has the same truth value in all models of $\mathbf{ZFC}_2$ with the same interpretation of the range of second-order quantifiers.

To make the state of the dialectic precise, and given the difficulty of interpreting Kreisel, it is worth pulling out the key moving parts of our interpretation of Kreisel's presentation:

> (**Assumption of Informal Rigour**) A putatively non-algebraic mathematical discourse D determines a particular structure S when we are informally rigorous in employing the relevant concept C corresponding to D, and this informal rigour can be manifested in a categorical axiomatisation $\mathbf{T}$ of C such that for any systems $\mathfrak{M}$ and $\mathfrak{M}'$ exemplifying S, both $\mathfrak{M}$ and $\mathfrak{M}'$ satisfy $\mathbf{T}$ and are isomorphic.
>
> (**Manifestation Thesis**) We become informally rigorous about a concept C through either (a) developing our mathematical understanding of C by working with it in practice (i.e. top-down rigour), or (b) characterising it through already accepted resources (i.e. bottom-up rigour). In the case of concepts for particular structures, this understanding can then be *manifested* by a categorical axiomatisation $\mathbf{T}$. (In other words, the existence of a categorical axiomatisation is necessary for us to have informal rigour about a concept determining a particular structure.)
>
> (**Segment Particularity Thesis**) We are informally rigorous about the concept *cumulative type structure below the first inaccessible*, and this concept is axiomatised by the theory $\mathbf{ZFC}_2$ + "There are no inaccessible cardinals" and determines a particular structure.
>
> (**CH-Determinateness Thesis**) The concept *cumulative type structure* suffices to determine a truth value for CH.
>
> (**Difference Thesis**) The kind of independence exhibited by CH (relative to $\mathbf{ZFC}_2$) and PP (relative to the axioms of geometry) are of fundamentally different kinds.

In what follows, we shall take the Assumption of Informal Rigour as an assumption (though we'll discuss how to flesh it out in more detail). This is just because I'm interested in exploring the idea; it's clearly a very controversial assumption! We'll argue that the Segment Particularity Thesis and CH-Determinateness Thesis can be challenged. We'll then argue that the Manifestation Thesis suggests that our thought is perhaps best axiomatised by something weaker than $\mathbf{ZFC}_2$. We'll also argue that the Difference Thesis still holds true.

6.3 Three Interpretations of Informal Rigour

In the last section, we saw some theses that one might extract from Kreisel's paper on informal rigour. In this section, I'll present three ways of interpreting this process of informal rigour that will be important for later.

6.3.1 Isaacson's Kreisel

One way of interpreting the process of informal rigour has been proposed by Dan Isaacson (2011). There he seems to commit himself to the Assumption of Informal Rigour in the following passage:

> We achieve understanding of the notion of mathematical structure not by axiomatizing the notion but by reflecting on the development of mathematical practice by which particular mathematical structures come to be understood, the natural numbers, the Euclidean [plane], the real numbers, etc.
>
> How do we know that such structures exist? The question is likely to be construed in such a way that it is a bad question. There is nothing we can do to establish that particular mathematical structures exist apart from articulating a coherent conception of such a particular structure. (Isaacson, 2011, p. 29)

as well as the Manifestation Thesis:

> ... if the mathematical community at some stage in the development of mathematics has succeeded in becoming (informally) clear about a particular mathematical structure, this clarity can be made mathematically exact. Of course by the general theorems that establish first-order languages as incapable of characterizing infinite structures the mathematical specification of the structure about which we are clear will be in a higher-order language, usually by means of a full second-order language. Why must there be such a characterization? Answer: if the clarity is genuine, there must be a way to articulate it precisely. if there is no such way, the seeming clarity must be illusory. (Isaacson, 2011, p. 39)

However, his interpretation of these notions is decidedly *not* objectual in the platonistic sense of concerning mind-independent abstract objects:

> The basis of mathematics is conceptual and epistemological, not ontological, and understanding particular mathematical structures is prior to axiomatic characterization. When such a resulting axiomatization is categorical, a particular mathematical structure is *established*. Particular mathematical structures are not mathematical objects. They are characterizations. (Isaacson, 2011, p. 38, my emphasis)

So, for Isaacson, the process of informal rigour can be understood as a *mind-dependent* activity in some sense. The process of informal rigour should not be understood as one where we pick out some pre-existing ontological objects, but rather as the determination of a particular structure using our thought and language, one that does not exist in advance of our characterising activity (in this sense, his view is quasi-idealist). This precision *in our concept* is then manifested by a categorical axiomatisation **T**.

Isaacson's claim that particular structures just *are* characterisations is a little puzzling; the claim that particular structures are literally numerically identical with theories (i.e. characterisations) has the whiff of a category mistake about it. However, it serves to show further how we might think of informal rigour as a process of mathematical claims being dependent upon our epistemological and conceptual activity, rather than any independently existing structural domain.

Isaacson's version of informal rigour does not commit him to an 'anything goes' version of conventionalism. First, given some employed concepts about which we are informally rigorous, there can be objective facts about what follows from that

concept.[14] This is visible from Isaacson's endorsement of the CH-Determinateness Thesis.[15] Moreover, we are able to fix infinitely many structures in this way via bottom-up characterisations.[16] For example, the categoricity of the natural numbers establishes infinitely many particular structures, e.g. the structure exemplified by $(n, <)$ for any chosen n. Since it is unclear whether or not Kreisel would have accepted Isaacson's interpretation, I shall refer to a character I call 'Isaacson's Kreisel' as a proponent of this view of informal rigour.

6.3.2 Weak Kreiselian Platonism

Isaacson's Kreisel represents a version of informal rigour which feeds into a quite anthropocentric characterisation of the notion of structure. On his characterisation, informal rigour concerning the concepts employed in a discourse is constitutive of establishing the relevant structure in question.

Instead, we might have a more platonistic conception of informal rigour. One might rather hold that structures are mind-independent, and there are many abstract concepts we can employ in talking about those structures.

Given a discourse D and employment of a concept C_0 underlying this discourse, informal rigour on this picture consists of a successive narrowing down and improvement of the concept C_0.[17] If C_0 does not already determine some particular structure S, this may then necessitate moving to a sharper concept C_1 to underwrite D. Once we have become sufficiently informally rigorous about the concept underlying D (this might take several iterations of conceptual refinement) and have

[14] A good question, one we do not have space to address here, is how Isaacson's version of Kreisel relates to Ferreirós (2016)'s account of mathematics as *invention cum discovery*.

[15] He writes:

> ...the independence of the continuum hypothesis does not establish the existence of a multiplicity of set theories. in a sense made precise and established by the use of second-order logic, there is only one set theory of the continuum. it remains an open question whether in that set theory there is an infinite subset of the power set of the natural numbers that is not equinumerous with the whole power set. (Isaacson, 2011, pp. 48–49)

[16] He writes:

> While indeed there are up to any given moment of course only finitely many theorems establishing categorical characterizations of structures, e.g. of the natural numbers, the real and complex numbers, the Euclidean plane, the cumulative hierarchy of sets up to a particular ordinal, one such theorem may establish categorical characterization of infinitely many particular substructures. (Isaacson, 2011, p. 38)

[17] Of course, there may be more than one concept involved, in which case we might have to consider a concepts $C_0, \ldots, C_\alpha$ instead. I suppress this complication; nothing in my arguments hangs on there being just one concept or many.

pinned down some mind-independent particular structure S with some concept C_2, we are then able to provide our categorical axiomatisation **T** corresponding to C_2.[18]

In many ways, at a practical level, the Weak Kreiselian Platonist and Isaacson's Kreisel have much in common. They both think that mathematics depends in some way on us, the Weak Kreselian Platonist because the ways we refine our concepts are presumably dependent upon us (even though they may be constrained), and Isaacson's Kreisel because mathematical structures are determined by our activity. They differ in that the Kreselian Platonist thinks that the structures we talk about, and plausibly the concepts employ, are independent of us and informal rigour allows us to make a selection between them. Isaacson's Kreisel, on the other hand, thinks that the structures are determined by us, rather than discovered.

6.3.3 Strong Kreiselian Platonism

There is a stronger version of Kreiselian Platonism. The key additional assumption is the following:[19]

> (**Set-Theoretic Uniqueness**) There is one and only one correct concept C for discourse that is sufficiently 'set-like' (i.e. concerns extensional objects), and it is possible for us to have informal rigour about C. Informal rigour should be understood as a way of approximating C ever more closely.

So, for the Strong Kreiselian Platonist, it is not only the case that we may refine concepts in coming to be informally rigorous but also that we tend towards exactly one such way of filling out the concept in the case of set theory.

We then have three figures; Isaacson's Kreisel, the Weak Kreiselian Platonist, and the Strong Kreiselian Platonist. We shall argue that for Isaacson's Kreisel and the Weak Kreiselian Platonist, the status of the informal rigour of the universe of sets (and in particular the Continuum Hypothesis) is questionable. The Strong Kreiselian Platonist can hold on to the full informal rigour of the set concept up to a certain level, but we will argue that their position faces a quadrilemma.

---

[18] One question, that we shall leave as an open question at the end of the paper, is how we should understand this process of conceptual refinement. For example: Do the concepts stay the same, or do they change when we refine our concepts? For the purposes of discussing informal rigour and whether or not CH is determinate, I'm not sure this matters so much, but for the future development of set theory (and mathematics more generally) we might wonder how conceptual refinement figures in debates about, for example, the temporal continuity of subject matter in mathematics. I am grateful to Chris Scambler for many hours of interesting discussion here.

[19] I am grateful to Leon Horsten for suggesting this interpretation, and Daniel Kuby for some additional discussion led me to realise that I also needed to consider the weaker form of Kreiselian Platonism as discussed in the last subsection.

6.4 Structural Relativity

We are now at a point where we have said a little more about how we might fill out an account of informal rigour, and provided some possible philosophical interpretations of the notion. For the purposes of our arguments in Sect. 6.5 and interpreting our own set-theoretic discourse, it will be useful to set up the idea of *structural relativity*.

Structural relativity is the idea that the structure isolated by a particular piece of mathematical discourse is contingent upon the logic used to underwrite it. It is discussed explicitly by Resnik (1997):

> In thinking about formulating a theory of structures we must take into account a phenomenon I will call structural relativity, the structures we can discern and describe are a function of the background devices we have available for depicting structures ... This relativity arises whether we think of patterns and structures as a kind of mould, format, or stencil for producing instances, or as whatever remains invariant when we apply a certain kind of transformation, or as an equivalence class or type associated with some equivalence relation. The structures we recognize will be relative to our devices for specifying forms, or transformations or equivalence relations. (Resnik, 1997, p. 250)

The idea then for Resnik is that the kind of structures we can talk about can vary contingent upon the logical resources we employ. For the same mathematical discourse D, we might pick many different formal theories to underwrite it, and many different kinds of structure might be thereby isolated. For example, he writes:

> If we limit ourselves to describing structures as the models of various first-order schemata, then the types of structures we will define will be like the more coarse-grained ones frequently found in abstract algebra. Here one starts by defining a type of structure such as a group, a ring, or a lattice with the intention of allowing for many non-isomorphic examples of the same type. As a result most of our structural descriptions will fail to be categorical. On the other hand, using second-order schemata, we can formulate categorical descriptions of the structures studied by (second-order) number theory, Euclidean geometry and analysis, and categorical extensions of [$\mathbf{ZFC}_2$] that are considered powerful enough for most mathematical needs.
>
> Thus, depending upon our logical resources, we might introduce:
> The First-Order Natural Number Structure,
> The Second-Order Natural Number Structure,
> The First-Order Structure of the Reals,
> The Second-Order Structure of the Reals,
> and so on.
>
> By going to stronger logics we get more fine grained versions of the various structures. (Resnik, 1997, p. 252)

So, for example, we can consider our talk about natural numbers as either formalised in first-order Peano Arithmetic (**PA**), or in second-order Peano Arithmetic ($\mathbf{PA}_2$). The latter axiomatisation corresponds (given the full semantics) to the particular structure of the standard model of natural numbers, the former on the other hand is a general structure that is can be both instantiated by the standard model (where, presumably, $Con(\mathbf{PA})$ holds), but also can be instantiated by non-isomorphic non-standard models (where, for example, $\neg Con(\mathbf{PA})$ can hold).

The above passage is fairly indicative of what seems to be a (false) dichotomy underlying parts of the literature; we are presented with the choice between either using first-order resources (where almost nothing is categorical, only finite structures) or full second-order resources (where an enormous amount of our mathematical talk is fully categorical).[20] This dichotomy does not adequately reflect the fact that in mathematical logic we have a wide range of logics intermediate between first-order and second-order. The properties of these logics are well-understood,[21] and it is surprising that they have not been considered in detail in the context of structural relativity. This is not to say that authors (including Resnik) intend this false dichotomy, just that largely speaking in the structuralist literature these are the two options proffered.

Admitting intermediate logics into interpretations of structural relativity opens up a host of possibilities. Once we free ourselves of the binary choice between first- and second-order resources, we have the option of considering many different formal theories for underwriting a discourse. There is a wide variety of options here, including increasing our resources beyond first-order with certain operators (e.g. ancestral logic) or alternatively allowing infinitary conjunctions or quantifier alternations. Since we will be interested here in theories that we can *use* in manifesting informal rigour, we set aside the use of infinitary resources. In the next section, we shall see how versions of set theory incorporating structural relativity given by weak second-order logic and quasi-weak second-order logic correspond to two natural positions about informal rigour concerning the cumulative hierarchy.

6.5 The Concept of Set, Degrees of Informal Rigour, and Structural Relativity

We are now in a position where:[22]

(1.) Informal rigour in the concepts underlying a discourse is manifested by axiomatisations that are categorical, either by top-down or bottom-up approaches.

[20] Isaacson, for example, writes:

> As Shapiro and others have long noted, the language in which to articulate our understanding of particular mathematical structures is second-order. . . (Isaacson, 2011, p.28)

[21] See, for example, Shapiro (1991, Ch. 9) or Shapiro (2001).

[22] This section, and in particular my discussion of what I'll call the Modal Definiteness Assumption, is enormously indebted to Chris Scambler. We worked on this together as part of a joint project, and I am very grateful for his kind permission to include the following discussion in this piece. Of course, any mistakes made in filling out the details should be attributed to me rather than Chris.

(2.) We have three different ways of interpreting informal rigour, via Isaacson's Kreisel, Weak Kreiselian Platonism, and Strong Kreiselian Platonism.
(3.) Structural relativity may come in to play, whereby the kinds of structures we isolate are contingent upon the background logic we use.

In this section, I'll consider some examples that show how we might not be fully informally rigorous about our set-theoretic discourse and set concept. I'll then argue that there are reasons to think that there may be a degree of structural relativity involved in the axiomatisation of our thought concerning sets. Nonetheless, I shall argue that we are (and have been) *partially* informally rigorous, and our discourse about *portions* of the hierarchy can be understood as about particular structures. To do this, I'll look at a Predicativism proposed by Feferman and Hellman, and then historically at Mirimanoff's thought concerning the Axiom of Foundation, before considering our own axiomatisation of set theory in terms of **ZFC** and our possible attitudes to CH.

In order to make out my conclusions, it will be useful first to analyse in a little more detail what we might expect from an account of informal rigour. Important for Kreisel's notion is that our concept of set, and the informal rigour we have about it, is a *source* for axioms. He writes:

> What one means here is that the intuitive notion of the cumulative type structure provides a coherent *source* of axioms; our understanding is sufficient to avoid an endless string of ambiguities to be resolved by further basic distinctions...[23] (Kreisel, 1967, p. 144)

Isaacson agrees, at least insofar as interpretation of Kreisel goes:

> In order actually to solve the continuum problem a formalizable derivation from axioms, of the kind which Cohen and Gödel's results show not to exist from the first-order axioms of **ZF**, must be found. This means that new axioms are required. (Isaacson, 2011, p. 16)

My point is the following: If we are informally rigorous about a discourse D and the concepts underlying it, and hence have determined a particular structure, we can expect the use of these concepts as a "coherent source of axioms" not to lead us in radically different directions. Of course, it is possible to have beliefs about a structure that turn out to be false (as when I believe an eventually false conjecture), but it should not be the case that radically different concepts, with radically different theories and consequences are legitimate ways of refining our current concepts. We therefore identify the following:

> (**The Modal Definiteness Assumption** or MDA) If we are informally rigorous about a mathematical discourse D, using a concept C_0 to determine a particular structure S, then there should not be two (or more) legitimate ways of refining C_0 (to some C_1 and C_2)

[23] Kreisel continues:"...like the distinction above between abstract properties and sets of something.", speaking about the distinction between intensional entities and sets (this intensionality he seems to diagnose as the source of the class-theoretic paradoxes). Since this diagnosis is rather controversial, I'll set it aside here.

such that C_1 motivates a theory $\mathbf{T}_1$ and C_2 motivates a theory $\mathbf{T}_2$ such that $\mathbf{T}_1$ and $\mathbf{T}_2$ are inconsistent with one another.[24]

Why do I call this assumption 'modal'? Throughout the rest of the paper, we will consider small perturbations concerning how things might have gone in the past, or might go in the future, and show that given these assumptions an agent's concepts might be expanded in different ways to incompatible extensions. This then casts doubt on the claim that their concept is informally rigorous and determines a particular structure regarding some subject matter.

The Modal Definiteness Assumption is definitely controversial, but also intuitively plausible. If our discourse and concepts already determine a particular structure (via informal rigour) then there should not be equally legitimate ways of sharpening our concepts that are inconsistent with one another, since the truth values of all claims in the discourse are already set by this structure. Therefore one of the two theories has to be false, thus one of the two concept-schemes is inferior, and so they are not equally legitimate.[25] Of course, what constitutes a 'legitimate' extension is going to be something of debate, but the rough idea is that a change or refinement of a concept is one that still coheres with the original, but adds well-motivated content. Whilst these are difficult ideas to make precise, I hope that examination of the examples I provide from the philosophy of set theory will make it clear that there may be such sharpenings, and hence by the MDA we may not be informally rigorous about our concept of set. However, let us first see how the MDA might play out in a positive case where we *do* take ourselves to have informal rigour.

6.5.1 The Radical Relativist

Suppose we believe that our discourse about the natural numbers, underwritten by our concept of natural number, is informally rigorous and this informal rigour is manifested by $\mathbf{PA}_2$ and the attendant Dedekind-categoricity theorem. Along comes the Radical Relativist who says to us: You cannot be informally rigorous about arithmetic, since there are legitimate consistent extensions $\mathbf{PA}_2 + Con(\mathbf{PA}_2)$ and $\mathbf{PA}_2 + \neg Con(\mathbf{PA_2})$ of $\mathbf{PA}_2$ that are inconsistent with one another (where $Con(\mathbf{PA}_2)$ is the consistency sentence for $\mathbf{PA}_2$ given the syntactic deduction relation for second-order logic[26]). What should are reaction be?

[24] Many thanks to Daniela Schuster for pressing me to become clearer about my formulation of the MDA.

[25] If you're familiar with debates in the philosophy of set theory, you might already see where I'm going here.

[26] This will, of course, not be complete. Nonetheless one can define this relation, see Button and Walsh (2018).

Our response should be the following: Of course these extensions are *formally* consistent, in the sense that assuming $\mathbf{PA}_2$ is ω-consistent (given the incomplete syntactic deduction relation for second-order logic) a contradiction is not derivable in either $\mathbf{PA}_2 + Con(\mathbf{PA}_2)$ or $\mathbf{PA}_2 + \neg Con(\mathbf{PA}_2)$. One is nonetheless clearly legitimate where the other is not. In particular, $\mathbf{PA}_2 + \neg Con(\mathbf{PA}_2)$ can only be true in models that are non-standard, both in that (i) the interpretation of the second-order variables has to be given by a Henkin semantics that permits a two-sorted first-order characterisation, and (ii) the theory also has consequences (assuming $\mathbf{PA}_2$ is in fact consistent) that do not accord with our concept of natural number, for example models of the theory contain a natural number n^*, such that for any particular standard natural number n given to me, n^* is greater than n. So it is simply not true that $\mathbf{PA}_2 + Con(\mathbf{PA}_2)$ and $\mathbf{PA}_2 + \neg Con(\mathbf{PA}_2)$ are both legitimate extensions of $\mathbf{PA}_2$, at least insofar as axiomatising our concepts and thought concerning the particular structure of natural numbers is concerned.

Moreover, there is no categoricity theorem for the theory $\mathbf{PA}_2 + \neg Con(\mathbf{PA}_2)$, and indeed it can have highly non-isomorphic models. In fact, since we must allow non-full Henkin interpretations here, we are effectively working in a two-sorted first-order framework, and so the usual trappings of first-order logic apply. So there can be no categoricity theorem for this theory, and hence no informal rigour.

This will provide a contrast case for our main examples; considering a Predicativism proposed by Feferman and Hellman, examining the historical situation with respect to Miramanoff and the Axiom of Foundation, and our contemporary situation with respect to set theory and CH.

6.5.2 The Predicative Iterabilist

We now consider a slightly different situation, one in which we have agents whose thought is best axiomatised by a version of set theory intermediate between first and second-order **ZFC**.

Suppose that one accepts that we are informally rigorous about the concept of natural number, but has extreme reservations about the whole of set theory. A view providing a predicative foundation for arithmetic has been advanced by Feferman and Hellman in a pair of papers Feferman and Hellman (1995) and Hellman and Feferman (2000).[27] In Feferman and Hellman (1995) they define a system **EFSC** (for **E**lementary theory of **F**inite **S**ets and **C**lasses) and provide a categoricity proof for natural number systems within **EFSC**. Suppose further that a Predicativist of the Feferman-Hellman variety expands their concepts and accepts the iterative conception as a conceptual *idea*, and hence regards **ZFC** as a (probably) consistent theory worthy of study, but has extreme reservations about informal

[27] I am grateful to Geoffrey Hellman for pointing to the position of Feferman and Hellman as a possible case study.

rigour concerning the notions of arbitrary subset and arbitrary well-order. Instead, they think that we can only be informally rigorous about things that are *predicatively* defined, and think that it's possible that our thinking might not be informally rigorous and fail to determine particular structures at large infinite ordinals. Call this character the *Predicative Iterabilist*. What should the Predicative Iterabilist say about our set-theoretic thought concerning the iterative conception?

To make our points (here and later) we first need to set up some terminology. Two background logics will be of special interest for us:[28]

Definition *Weak second-order logic* is the logic in which we allow the same vocabulary as second-order logic $\mathscr{L}^2_K$ (where K are the non-logical symbols) but with function variables removed.

Its semantics is given by letting the second-order quantifiers range over *finite* relations. Let $\mathfrak{M}$ be a model with domain M. We define a *finite assignment s on* $\mathfrak{M}$ as assignment s that assigns a member of M to each first-order variable, and a finite n-place relation on M to each n-place relation variable. Satisfaction is defined in the usual manner for the first-order connectives and quantifiers, and second-order quantification is handled by the clause:

> $\mathfrak{M}, s \models \forall X\phi$ iff for every finite assignment s' that agrees with s (except possibly at X), $\mathfrak{M}, s' \models \phi$.

The instances of Comprehension $\exists X\forall y(X(y) \leftrightarrow \phi(y))$ which are valid on a structure $\mathfrak{M}$ are those where the extension of ϕ is finite in $\mathfrak{M}$.

Let $\mathbf{ZFC}_{2W}$ be set theory formulated in weak second-order logic with instances of the replacement scheme for each formula of the weak second-order language.

Definition *Quasi-Weak Second-Order Logic* is the same as Weak Second-Order Logic, but in the semantics each variable assignment assigns countable relations to the variables (i.e. we assign countable relations instead of finite ones). So $\forall X\phi$ holds iff for all countable X, ϕ holds.

Let $\mathbf{ZFC}_{2QW}$ be set theory formulated in quasi-weak second-order logic with instances of the replacement scheme for each formula of the quasi-weak second-order language.

It is useful to identify some facts off the bat:[29]

Fact Both $\mathbf{ZFC}_{2QW}$ and $\mathbf{ZFC}_{2W}$ are able to characterise categorically the natural numbers (i.e. any two models of $\mathbf{ZFC}_{2QW}$ and $\mathbf{ZFC}_{2W}$ always have the standard natural numbers as their standard model of arithmetic, and indeed any two models of $\mathbf{PA}_2$ with the full semantics within a model of $\mathbf{ZFC}_{2QW}$ or $\mathbf{ZFC}_{2W}$ are isomorphic).

[28] The presentations given here are heavily indebted to Shapiro (2001).

[29] See Shapiro (2001) for discussion of these results.

This is because we can characterise the notion of finiteness in both quasi-weak and weak second-order logic.[30] The same goes for the rational numbers.[31]

Fact $\mathbf{ZFC}_{2QW}$ is able to characterise the theory of real analysis up to isomorphism. Essentially, this is because we can characterise the completeness principle for the reals in $\mathbf{ZFC}_{2QW}$.[32] In $\mathbf{ZFC}_{2W}$, however, one cannot characterise the reals up to isomorphism, since the Löwenheim number of Weak Second-Order Logic is $\aleph_0$.[33]

Fact $\mathbf{ZFC}_{2QW}$ is able to characterise the notion of well-foundedness, that is, all models of $\mathbf{ZFC}_{2QW}$ are well-founded.[34]

Fact $\mathbf{ZFC}_{2W}$ is *not* able to characterise the notion of well-foundedness (i.e. there are models of $\mathbf{ZFC}_{2W}$ with a non-well-founded membership relation).[35]

These facts show that quasi-weak second-order logic has substantially more *expressive power* than weak second-order logic; we can characterise more notions within in it (and in turn, the versions of set theory formulated in the respective logics differ in their expressive power and intended models).

So, we have several logics and versions of **ZFC**-like set theory rendered in them in view. Now, the Predicative Iterabilist will hold that we are informally rigorous about the natural numbers, but have grave worries about our informal rigour concerning the iterative conception in general. In this case, we might think that our thought about **ZFC**-based set theory and the concept of cumulative type structure is best axiomatised by $\mathbf{ZFC}_{2W}$. There we are able to identify the rational and natural numbers up to isomorphism, but the real numbers cannot be so identified, and various large well-orderings (e.g. ω_1^{CK}) cannot be characterised up to isomorphism.[36]

If you are a Predicative Iterabilist, you are thus likely to hold that our talk about the concept *cumulative type structure* is only *partially* informally rigorous, and this level of partial informal rigour is manifested in $\mathbf{ZFC}_{2W}$. We thus have a coherent position on which a level of informal rigour is manifested in a logic stronger than first-order but weaker than second-order.

[30] See Shapiro (2001), p. 161, and Theorem 16 and Corollary 17 on p. 162.

[31] This is because we can characterise the notion of *minimal closure* in the two logics, and the rational numbers can be characterised up to isomorphism as an infinite field arising from the minimal closure of {1} under the field operations and their inverses. See Shapiro (2001, p. 161).

[32] See Shapiro (1991, pp. 164–165).

[33] See Shapiro (2001, pp. 161–162).

[34] Assuming Choice in the meta-theory, the fact that every countable class is a set in a model of $\mathbf{ZFC}_{2QW}$ ensures this. See Shapiro (1991, p. 165).

[35] This is because there is a natural equivalence between being a model of $\mathbf{ZFC}_{2W}$ and being an ω-model of **ZFC** (see Shapiro 1991, p. 162, Corollary 17) and there are non-well-founded ω-models of **ZFC**.

[36] See here Shapiro (1991, p. 163).

We can make out this point using the MDA.[37] If I am a Predicative Iterabilist I believe I have grounds for the determinacy of thought concerning the natural numbers, but not the full real numbers, impredicatively defined. How can they make the grounds for this indeterminacy precise using the MDA? Well, they accept the use of $\mathbf{ZFC}_{2W}$ as underwriting our theory of sets by recognising as absolute the finite sets of a given set, and the natural numbers as determinate. This framework supports informal rigour regarding the concept *finite subset of the natural numbers*, and this is enough to pin down the natural numbers up to isomorphism. But when we look to expand our theory to the real numbers this framework can be extended in two different incompatible ways. On the one hand, we can extend our determinate theory of the natural numbers to the classical continuum via Dedekind-cuts or equivalence classes of Cauchy sequences in the rationals. On the other hand, we could extend to the intuitionistic continuum, as developed by Brouwer, Heyting and others. These two extensions formally contradict one another; for example the intuitionistic theory proves that all functions are (uniformly) continuous, whereas in the classical continuum we have many discontinuous functions.[38] Thus, the Predicative Iterabilist can spell out why she does not think that there is informal rigour concerning the reals in terms of the MDA. Moreover, if we restrict the discussion to the *classical* continuum, there are still different ways of extending her theory $\mathbf{ZFC}_{2W}$; we might choose to include or exclude axioms of definable determinacy. She also has a quick explanation of why the MDA does not speak against her belief in determinateness concerning the natural numbers; there are no known legitimate expansions of her concept of natural numbers that motivate inconsistent theories. As mentioned in Sect. 6.5.1, our concept of natural number clearly excludes known independent sentences (like Gödelian diagonal sentences) as being theory expansions concerning a legitimate conceptual refinement.

Later (Sect. 6.5.4) we shall see that a similar argument can be made for the believer that the reals are determinate, assuming that we accept axioms of definable determinacy (axioms with close relationships to large cardinals). Of course, the believer in the Segment Particularity Thesis on the basis of the quasi-categoricity of $\mathbf{ZFC}_2$ will reject this application of the MDA. We will discuss the place of the quasi-categoricity theorem later (Sect. 6.6), for now we just note that the example as presented shows that we can have a coherent position on which our reasoning is axiomatised by a set theory couched in a logic intermediate between first- and second-order and this belief can be made precise on grounds involving the MDA. Before we discuss CH, we will mention a historical example.

[37] I am grateful to Geoffrey Hellman for suggesting this as a possible objection to my final position that ZFC_{2QW} is a plausible candidate to underwrite our discourse involving sets. By re-purposing the objection to the case of the Predicative Iterabilist, I think that it bolsters the role of the MDA in making precise grounds for indeterminacy.

[38] Examples can be multiplied. A simpler example is the intuitionistic theorem that it is not the case that any given infinite sequence of 0s and 1s, the sequence is either composed of 0s everywhere or contains a 1 somewhere, contradicting the obvious classical fact. See Dummett (1977, Ch. 3), for a proof.

6.5.3 *Miramanoff's Informal Rigour*

The following example will provide an example where we have a failure of informal rigour on the basis of the MDA, but might nonetheless think that substantial parts of mathematics are informally rigorous, and as such we have partial informal rigour in the notion of set. We'll see, however, that the example is more analogous to PP than CH (the latter we consider in Sect. 6.5.4).

In 1917, Dimitry Mirimanoff wrote a paper entitled 'Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles'. In this paper, he considers Russell's Paradox and the Burali-Forti Paradox, and identifies two kinds of sets; the 'ordinary' ones and the 'extraordinary' ones. These were to be differentiated by whether or not they contain infinite descending sequences of membership; the ordinary ones do not (in current terminology: they have a well-founded membership relation) and the extraordinary ones do (in current terminology: they have a non-well-founded membership relation):

> I will say that a set is *ordinary* just in case it gives rise to finite descents, I will say that it is *extraordinary* when among its descents are some that are infinite. (Mirimanoff 1917, p. 42, my translation)[39]

It is clear that Mirimanoff (1917) was undecided about whether the Axiom of Foundation was a basic principle about sets. It is also fairly clear, we think, that he was *not* fully informally rigorous about set theory. To see this, it suffices to consider what theory might have underwritten his thinking about sets, and show that there are different legitimate extensions that are inconsistent with one another.

Clearly Mirimanoff thought that sets were extensional and he explicitly discusses the axioms of pairing and union, as well as replacement. For the purposes of our discussion, let us assume that he was clear that his notion of set supported at least the first-order axioms of **ZF** without the Axiom of Foundation. (It doesn't matter so much whether or not these were *actually* Mirimanoff's views, as long as this character is at least possible it shows the kinds of situations that are compatible with informal rigour in set theory.)

Can Mirimanoff's level of informal rigour support more? Is he informally rigorous about the Axiom of Foundation? We answer this negatively using the Modal Definiteness Assumption. We argue that there are legitimate extensions of Mirimanoff's concept that support inconsistent theories of sets (such as **ZF** and **ZF**-Foundation+AFA).[40] Clearly the former is a legitimate extension, since it is

[39] The original French reads:

> Je dirai qu'un ensemble est *ordinaire* lorsqu'il ne donne lieu qu'a des descentes finies; je dirai qu'il est *extraordinaire* lorsque parmi ses descentes il y en a qui sont infinies. (Mirimanoff, 1917, p. 42)

[40] Here AFA denotes Aczel's Anti-Foundation Axiom, which has strong affinities with the graph conception of set. See Aczel (1988).

what we (as a matter of fact) use now on the basis of our concept of cumulative type structure. Is the latter a legitimate extension of **ZF**-Foundation? One might be tempted to answer no: The iterative conception of set clearly prohibits the existence of non-well-founded sets.

The iterative conception is emphatically *not* Mirimanoff's conception of set, however. Whilst he has the concept of ordinal and rank in play,[41] it is not really until Zermelo (1930) that we start to see the idea of cumulative type structure emerge, solidified in Gödel's work on L (in Gödel 1940), and it was not until the late 1960s and 1970s that the idea of the iterative conception and its relation to **ZFC** were fully isolated.[42] Indeed, Mirimanoff seems to treat non-well-founded sets as legitimate objects worthy of study, formulating a specific notion of isomorphism known as *tree-isomorphism* that works for both non-well-founded and well-founded sets.[43] The following situation is then possible: Suppose that instead of the iterative conception becoming the default conception of set, the graph conception of set (on which sets are viewed as given by directed graphs) became the default set-theoretic conception. We might, for example, have been persuaded by considerations about non-well-founded sets emerging in computer science (as when they are used to model concurrent processes).[44] Then, it seems reasonable to accept that Mirimanoff's intellectual descendants would have accepted that there were non-well-founded sets. By the MDA, he can't then have been fully informally rigorous, since there are inconsistent ways of extending the concept he was employing about his discourse.

It is then tempting to say that Mirimanoff's thinking might be best captured by *first-order* **ZF** without Foundation. We should resist this temptation. Mirimanoff's context is plausibly one in which he was informally rigorous about what the natural numbers were, and indeed his work comes after Dedekind's categoricity proof (in Dedekind 1888). In particular, his definition of well-foundedness depends on the notion of finiteness; he characterises well-founded sets as those which only have finite descending membership chains, rather than using the contemporary first-order statement of the Axiom of Foundation in terms of the claim that every non-empty set A contains a set B such that $A \cap B$ is empty (a formulation which appears in Zermelo 1930).[45] But, by the Compactness Theorem, finiteness cannot be characterised using first-order logic, nor can the natural numbers.[46] It is overwhelmingly likely

[41] The notion of ordinal recurs throughout his discussion of the Burali-Forti Paradox, and he discusses the notion of rank on p. 51 of Mirimanoff (1917).

[42] In Boolos (1971), for example. See Kanamori (1996) for a thorough discussion of the history.

[43] See Aczel (1988, p. 105).

[44] See here Incurvati (2014) for a description of the graph conception and Aczel (1988) for a summary of non-well-founded sets (as well as some useful historical remarks in Appendix A).

[45] See here Aczel (1988, p. 107). Independently, von Neumann presented this formulation in 1929.

[46] In fact, being able to capture these two notions is roughly equivalent, since "x is finite" can be parsed in terms of being bijective with a standard natural number, and "x is standard natural number" can be parsed as being a finite successor-distance away from 0. See Shapiro (2001, p. 155) for the details.

that he would have not accepted non-standard models of arithmetic as legitimate interpretations in the same sense as his own.

Since Mirimanoff was also well aware that arithmetic could be coded in set theory, we are at a point where we would like to say that his discourse about parts of set theory such as the natural numbers and finite sets *are* informally rigorous and determined a particular structure. It is also plausible (putting aside worries of Predicativism) that he was informally rigorous around 1915 about the notion of real number, by this stage he was working on the intellectual foundations that had already been laid by Cauchy, Weierstrass, Cantor, and Dedekind, and the categoricity of the real line had been proved. However, by the MDA, his discourse about set theory in general was not informally rigorous. Thus, if we are to provide an axiomatisation for underwriting his discourse and concept of set, we should use a theory and logic that is not fully categorical, but nonetheless can identify parts of set theory up to isomorphism.

What should we say about Mirimanoff's level of informal rigour? Well, to review:

(1.) His concept of set did not clearly support the Axiom of Foundation.
(2.) It is highly plausible that he was informally rigorous about the natural numbers and the real numbers.
(3.) It is highly plausible that he was informally rigorous about the concept of well-order (being able to distinguish and talk about the extraordinary and ordinary sets).

We can then say that Mirimanoff's level of informal rigour about set theory can be roughly characterised by $\mathbf{ZFC}_{2QW}$-Foundation (i.e. $\mathbf{ZFC}_{2QW}$ with the Axiom of Foundation removed). There, we can characterise the usual objects of mathematics including the real, rational, and natural numbers (since the categorical characterisations of these theories do not depend on the Axiom of Foundation). Moreover, he can formulate and discuss his worries about well-foundedness in this logic. However, he is not fully informally rigorous, since there are incompatible legitimate expansions of the concept he was working with (namely to one supporting the foundation axiom and to one supporting its negation).

We should remark though that Mirimanoff's situation is more like the situation we have with the Axiom of Parallels in geometry, rather than what we have in $\mathbf{ZFC}_2$ with respect to CH. This is because there is no categoricity proof for $\mathbf{ZFC}_2$-Foundation as there are models of $\mathbf{ZFC}_2$-Foundation in which Foundation holds and others in which it fails. So whilst our example shows that there might have been a case where we failed to be informally rigorous about our notion of set, it does not yet show the possibility of a situation where *we* are not, where *we* have the iterative conception of set.

6.5.4 Modal Definiteness and the Continuum Hypothesis

So then: What now about our own thought concerning the Continuum Hypothesis? My contention is that, given the Modal Definiteness Assumption, we have good reason to think that we are *not* fully informally rigorous about our concept of set. To see this, it is useful to consider two active programs targeting the resolution of CH in the contemporary foundations of set theory, namely *forcing axioms* and Woodin's *Ultimate-L* programme.

We omit the details here, since they are technically rather tricky, and many questions are still open. A rough description of each, however, will help to see the senses in which they present legitimate conceptual refinements of our concept of *cumulative type structure*. Both kinds of programme attempt to capture notions of 'maximality' in some way. Ultimate-L does so by incorporating large cardinals in an elegant manner, potentially providing a model in which many questions are decidable but large cardinals can also exist.[47] Forcing axioms on the other hand ensure that various kinds of subset exist; in technical terms, they assert that the universe has already been saturated under the existence of generic filters for certain partial orders and families of dense sets. Both represent somewhat different takes on how our concept of set may develop; Ultimate-L focusses on the development of large cardinals and inner model theory, whereas forcing axioms try to capture the idea of a rich process of subset formation.

Crucially, if we take the Ultimate-L approach, we can prove CH, and strong forcing axioms such as the Proper Forcing Axiom (PFA) imply ¬CH. They therefore represent inconsistent extensions of our current best theory of sets. They also both seem legitimate; both correspond to natural ways we might develop our set concept.

Given the MDA, it seems then that we are not fully informally rigorous about our concept of set. It is also plausible, however, that we have a good deal of informal rigour. We seem to have informal rigour about the natural numbers, where the only known independent statements are all equivalent to consistency statements, and the negation of these are illegitimate extensions (assuming that we think the axioms really are consistent). For second-order arithmetic, under both Ultimate-L and PFA there are no obvious analogues of CH; both programmes imply Projective Determinacy and there are no known sentences of **ZFC** independent from the theory **ZFC**−Powerset+$V = H(\omega_1)$ (other than Gödelian-style diagonal sentences). It also seems clear that our concept of cumulative hierarchy supports the idea that we are informally rigorous about the claim that all sets are well-founded.

[47] Whether we can construct Ultimate-L depends crucially on several conjectures in inner model theory. See Woodin (2017) for details. The key point is that if we are able to build a model that is 'L-like' and contains a supercompact cardinal, such a model would be able to tolerate all known large cardinal axioms that are also consistent, in contrast to the situation with $V = L$ and measurable cardinals (assuming that the existence of a measurable cardinal is, in fact, consistent).

Given this, it seems that our level of informal rigour in the cumulative hierarchy of sets might be top-down manifested by $\mathbf{ZFC}_{2QW}$. In quasi-weak second-order logic (and hence $\mathbf{ZFC}_{2QW}$) one can (bottom-up):[48]

(1.) Characterise $H(\omega_1)$ up to isomorphism by the theory consisting of:

(i) Extensionality
(ii) The axiom "Every set is countable".
(iii) The Axiom of Foundation, expressed as the claim that there is no ω-length infinite descending $\in$-sequence.
(iv) The sentence in quasi-weak second-order logic expressing "every countable subclass of the domain of discourse is the extension of a set".

This further bolsters our earlier claim that $\mathbf{ZFC}_{2QW}$ underlies our set-theoretic thought, since (given Projective Determinacy) no MDA-style argument is forthcoming for $H(\omega_1)$.

(2.) The field of reals $(\mathbb{R}, +, \times, <)$ is the only model (up to isomorphism) of the theory of ordered fields with the sentence of quasi-weak second-order logic expressing the claim that all Cauchy sequences converge and the Archimedean property that for every x in the domain of discourse, there is a finite sequence $\langle y_i | 0 \leq i \leq n \rangle$ of elements of the domain such that $x < y_n$, $y_0 = 1$, and for all $i < n$, $y_{i+1} = 1 + y_i$.

(3.) The standard model of second-order arithmetic can be characterised up to isomorphism (since every subset of natural numbers is countable, and quasi-weak second-order logic has an absolute interpretation for the range of the variables concerning countable relations).

However we can also point out:

Fact There are models of $\mathbf{ZFC}_{2QW}$ in which CH holds, and models of $\mathbf{ZFC}_{2QW}$ in which CH fails.[49]

[48] I am grateful to an anonymous reviewer for suggesting the specifics of these examples.

[49] I am grateful to Victoria Gitman for working with me on the following proof:

Proof Start in a model $\mathfrak{M} \models \mathbf{ZFC} + \neg\mathsf{CH}$ (by preparatory forcing if necessary). Next collapse $|\mathcal{P}^{\mathfrak{M}}(\omega)|$ to ω_1 using the forcing poset $Col(\omega_1, \mathcal{P}(\omega))$ in $\mathfrak{M}$. By design, $\mathfrak{M}[G] \models \mathsf{CH}$. But $\mathfrak{M}[G]$ also has the same countable relations on members of $\mathfrak{M}$ as $\mathfrak{M}$ itself, since it is a standard fact about $Col(\omega_1, \mathcal{P}(\omega))$ that it is countably closed. (If a countable relation R were added, one can look at the countably many conditions $p_n \in Col(\omega_1, \mathcal{P}(\omega))$ forcing that $\dot{x} \in \dot{R}$, and (by countable closure) infer that R was already in $\mathfrak{M}$.) Thus $\mathfrak{M}$ and $\mathfrak{M}[G]$:

(i) Have the same countable relations on sets in $\mathfrak{M}$ (for this reason $\mathfrak{M}$ and $\mathfrak{M}[G]$ have the same reals).
(ii) Differ on the truth value of CH.

Hence $\mathfrak{M}[G]$ thinks that both $\mathfrak{M}$ and $\mathfrak{M}[G]$ satisfy $\mathbf{ZFC}_{2QW}$ (since, according to $\mathfrak{M}[G]$, $\mathfrak{M}$ has all its countable relations) but differ on CH. Hence CH is not fixed by $\mathbf{ZFC}_{2QW}$. □

Thus, given the MDA and the Manifestation Thesis, we might think that our current level of informal rigour is manifested by $\mathbf{ZFC}_{2QW}$; a logic intermediate between first- and second-order. In this theory, when we can construct an argument for indeterminacy from the MDA we do not have the ability to provide a categorical characterisation, but if no such MDA-style argument is forthcoming (as is the case for the reals under Projective Determinacy) we *can* characterise many of the relevant structures up to isomorphism.

6.6 Objections and Replies

In this section I'll consider some objections and replies. These will not only help to shore up my position, but also will help to see some features of the account.

Objection *What about the Zermelo Categoricity Theorem?* One question for the arguments I have posed is immediate: What becomes of the Zermelo Quasi-Categoricity Theorem? One might think that the theorem shows that our thought about the sets is informally rigorous and determines some particular structures (for example those with a specific number of inaccessible cardinals). Earlier, I claimed that it is plausible that there are extensions of our current set concept that support Ultimate-L and others that support forcing axioms (let's take PFA to make things concrete). I then claimed on the basis of the MDA that our set-theoretic discourse and concepts were not informally rigorous. But this is not so (so one might counter-argue) whilst both PFA and Ultimate-L are (let's assume) *syntactically* consistent with $\mathbf{ZFC}_2$, only one of them can be true under $\mathbf{ZFC}_2$ with the full semantics, the other will require a Henkin-style interpretation to make both it and $\mathbf{ZFC}_2$ true. So it is just not correct to say that both are legitimate; the concept that motivates a theory that is false under the full semantics requires non-standardness of a certain kind (albeit not as serious as the one required for e.g. $\neg Con(\mathbf{ZFC}_2)$).

The issue here is that this objection assumes that we have access to the range of the second-order variables in making the criticism. We *already* need to be informally rigorous about the range of second-order variables if we are to hold that $\mathbf{ZFC}_2$ is a good encoding of our level of informal rigour. Similar points have been repeatedly stressed throughout the literature,[50] but it is particularly relevant to the current context; a categoricity theorem is meant to encode informal rigour that we have about a certain subject matter, not *give* us informal rigour (unless we have *already*

[50] See Meadows (2013) for a survey. Hamkins is also explicit about the point when discussing a version of the categoricity argument in Martin (2001):

> The 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? (Hamkins, 2012, p. 427)

accepted some resources for a bottom-up characterisation). If we don't have full informal rigour about set theory (which I've argued for on the basis of the Modal Definiteness Assumption) it is not necessary for us to accept that the categoricity theorem yields genuine clarity.[51]

It is instructive here to consider our different interpretations of informal rigour. Isaacson's Kreisel should accept (contrary to what Isaacson claims) that there are different legitimate extensions of our concept of set. This is because for Isaacson's Kreisel, informal rigour is dependent upon the degree to which we have understood a mathematical subject matter. If we expand our concept of set C_0 to one C_1 producing a consistent axiomatisation (as, let's assume, both Ultimate-L and PFA do) our *understanding* should be cashed out in terms of this new concept C_1, and this determines (*given* that we are employing C_1) a subject matter that supports either PFA or Ultimate-L, depending on which route we pick. Given then that for Isaacson's Kreisel the subject matter we talk about is determined by the concepts we employ, he should accept that we are able to go in different possible directions with our concept, and thus that we are currently not informally rigorous; our set-theoretic discourse is ambiguous between several different sharpenings of the notion.

For exactly the same reason, the Weak Kreiselian Platonist should accept that we are not fully informally rigorous about our concept of set. Recall that for her, informal rigour should be understood as coming to employ ever more platonistically existing precise concepts of set. But for this reason, it's entirely possible that we select one concept that supports PFA in the future and also possible that we select one that supports Ultimate-L. In this way, our thinking might be currently *ambiguous* between several different sharpenings of the concept.

The only person who can argue that the quasi-categoricity theorem in fact shows that $\mathbf{ZFC}_2$ encodes our level of informal rigour is the Strong Kreiselian Platonist. They hold that there is a *unique* correct concept that we are tending towards using informal rigour. This concept can then serve to interpret the second-order variables, given that $\mathbf{ZFC}_2$ is already quasi-categorical. Therefore (they claim) the case as I've set things up is not possible; *one* of PFA and Ultimate-L (or neither) is correct about this concept, and the process of informal rigour will lead us towards it. Therefore, exactly one or neither of PFA and Ultimate-L is legitimate, and it is just not possible to legitimately expand our concept in incompatible ways and at least one of PFA and Ultimate-L demands a non-full Henkin semantics for its interpretation. Hence, even accepting the MDA we can have informal rigour; simply put there are *not* incompatible *legitimate* extensions of our concept.

[51] A different move here would be to shift to *internal* categoricity. If one buys the MDA, however, one will be forced to accept *some* indeterminacy, blocking the argument to determinacy of CH. For example, even given an internal categoricity argument, indeterminacy in either the range of the first-order quantifiers or in the use of classical logic blocks the argument, in the first case because an internal categoricity result only determines CH within some restricted first-order domain, and in the second case because the proof of categoricity itself uses classical logic. See Scambler (Under review) for discussion of this issue.

This represents a coherent position, but not one that I find very plausible due a quadrilemma that I'll develop over the next few pages. The Strong Kreiselian Platonist has to accept that we simply *could not* coherently follow a different intellectual path from the one we have. But this is an enormously strong claim! What about cases where the kinds of modelling requirements we encounter are very different? Suppose, for example, that there are two physically (or even metaphysically) possible worlds W_1 and W_2 at which the modelling requirements for foundations are very different, and W_1 suggests Ultimate-L where W_2 suggests PFA. Should we insist that the agents at those worlds with different modelling requirements are doing something illegitimate if they select the 'wrong' concept of set? It seems to me that the agents in the two different cases simply employ different concepts, and use them to talk about different subject matters. But the Strong Kreselian Platonist has to either (a) accept that there is a fundamentally 'correct' interpretation for the second-order variables, and the thinking of one of the two communities' thinking is quite simply flawed, or (b) has to deny that such a situation is really possible. I do not find (b) especially plausible since possible worlds are pretty easy to come by.[52]

The situation can be made more vivid by a kind of pessimistic probabilistic argument. Assume that we do have a fully determinate interpretation of $\mathbf{ZFC}_2$. Notice that it might be that in fact *both* Ultimate-L and PFA are false in their full generality, even if one is correct about the status of CH. In fact, there are myriad different ways we might develop our set-theoretic axiomatisation, so why should we expect the one *we* pick to be right? Our understanding of the *Generalised* Continuum Hypothesis tells us that we can consistently have pretty much whatever pattern we like for the cardinal behaviour of infinite powersets (not to mention a whole gamut of other set-theoretic principles). So, if we believe that there really is a fully determinate $\mathbf{ZFC}_2$ model below the first inaccessible, it is overwhelmingly unlikely (without further argument) that we pick *exactly* the right axiomatisation, and it is *we* who are saying false things, and can only be interpreted as speaking consistently about non-standard Henkin interpretations.

If, given the Strong Kreiselian Platonist's position, we can coherently justify false set-theoretic principles, we obtain the following further counter-intuitive consequence: We can come to be *less* precise about the structure we talk about by developing our concept of set and accepting new axioms. Presumably, the Strong Kreiselian Platonist will want to assert that, given an agent A that has come to accept some false axiom(s) $\phi_0, \ldots, \phi_n$ extending $\mathbf{ZFC}_2$ that can be satisfied in a transitive model, we should (given a principle of charity) interpret A as saying true things

[52] For example Ben-David et al. (2019) showed that a certain learnability problem in machine learning is equivalent to CH. Much of the discussion of this problem (e.g. in Taylor 2019) consists of whether or not the algorithms in question are 'real-world' implementable. But if we just have to find some possible world or other rather than the actual world, then these worries about implementation are not so concerning. We can then easily cook-up possible worlds (in some loose sense of possibility) such that in one the evidence points to the learnability of the problem and another in which it points in the other direction.

about the relevant transitive models in which $\phi_0, \ldots, \phi_n$ can be realised, with a Henkin semantics for the relevant second-order variables. This is all well and good when there is an obvious unique model that can be identified as the place to interpret what she says. For example, suppose that we are considering the concept *cumulative type structure below the first inaccessible*, and further that A believes $V = L$, but (as it turns out) there are non-constructible reals below the first inaccessible. Then, letting κ be the least inaccessible, $(L_\kappa, \in, \mathcal{P}^L(\kappa))$ is a natural Henkin model in which to interpret A's discourse, and A has not lost precision in developing their concept of set to one motivating $V = L$. However, if $\phi_0, \ldots, \phi_n$ imply that there are unboundedly many measurable cardinals (as many of the candidate extensions of $\mathbf{ZFC}_2$ do), then we can point to the following:

Fact If a theory $\mathbf{T}$ (extending $\mathbf{ZFC}_2$ and mentioning only set-many parameters) is such that $\mathbf{T} \vdash$"There are unboundedly many measurable cardinals", then there is no least model of $\mathbf{T}$ under inclusion (within any $(V_\alpha, \in, \mathcal{P}(V_\alpha)) \models \mathbf{ZFC}_2$).[53]

The core philosophical point is the following: Supposing that the advocate of PFA also accepts the existence of unboundedly-many measurable cardinals, if one of Ultimate-L and PFA+"There are unboundedly many measurable cardinals" is false, then there is no easily identifiable unique model in which the agent accepting the false theory can be interpreted. Thus, by accepting more axioms on the basis of conceptual refinement (and thus, one might think, becoming *more precise* about their concept of set) they *lose* precision concerning the structure they talk about, compared to when they do not accept a refined axiomatisation and stick with $\mathbf{ZFC}_2$. This, one might think, is undesirable; we should become more precise, not less precise, by refining our concepts (at least insofar as mathematics is concerned).

Geoffrey Hellman vigorously objects to the conclusion that any of these alternatives undermines the Strong Kreiselian Platonist's position. My arguments are not meant to be knock-down, and indeed one can dig in one's heels here. However, if one does so, one will have to take on one of the following horns of a quadrilemma. Either:

(i) We will not, as a matter of fact, go astray in justifying new axioms extending $\mathbf{ZFC}_2$.
Challenge: If we take this horn of the quadrilemma, we then have to explain *why* we will not go astray in justifying new axioms. This looks like a difficult task and has the whiff of mysticism about it.

[53] I thank Monroe Eskew for discussion of the following:

Proof Let $\mathfrak{M}$ be a transitive model of $\mathbf{T}$ and let $\alpha \in \mathfrak{M}$ be such that $\alpha > rank(a)$ for every parameter a mentioned in $\mathbf{T}$. Let κ be a measurable above α. Then the embedding induced by the measurability of κ produces a proper inner model $\mathfrak{N}$ of $\mathbf{T}$ within $\mathfrak{M}$ (after finding a suitable Henkin interpretation for the second-order variables). Repeating the process yields the conclusion that there is no model of $\mathbf{T}$ least under inclusion contained in $\mathfrak{M}$. If $\mathbf{T}$ does not contain parameters, then one measurable cardinal suffices. □

(ii) We accept $\mathbf{ZFC}_2$, but also hold that we cannot justify axioms extending it (except perhaps large cardinals). $\mathbf{ZFC}_2$ (possibly with large cardinals added) is the limit of our possible justifications.
Challenge: This option essentially gives up on trying to resolve any sentences that are not consequences of large cardinals (e.g. CH).

(iii) An agent can become *less* precise by refining their set concept (if they pick an axiom with some false consequences).
Challenge: This response seems counter-intuitive; conceptual refinements should result in more rather than less precision.

(iv) One rejects the principle of charity, and accepts that in coming to justify new axioms, we might just say false things about the structure of sets, rather than true things about a range of structures.
Challenge: This likely ascribes a widespread error theory to many attempts to extend $\mathbf{ZFC}_2$. In particular (given the pessimistic probabilistic argument) it is likely that *we* will be in error in futures in which we accept a theory resolving questions independent of $\mathbf{ZFC}_2$ (plus large cardinals).

It is not impossible to take on one of the horns of the quadrilemma. However, we should contrast the position offered by the Strong Kreiselian Platonist with the alternative offered by either Isaacson's Kreisel or the Weak Kreiselian Platonist. They can argue that whilst we are not *yet* informally rigorous about our concept of set, and statements like CH are indeterminate given the concept we employ, we might be informally rigorous in the future. Just look, for example, at the progress that has been made in the hundred years or so since Mirimanoff was writing; our concept of set now *clearly* underwrites the claim that all sets are well-founded. Perhaps in the future we will come to a fully informally rigorous conception of set on which MDA-style arguments are not possible. However, even in this case we should acknowledge that it is not the case that things *had* to be this way. Both the Weak Kreiselian Platonist and Isaacson's Kreisel can avoid each of the problems for the Strong Kreiselian Platonist by accepting that what we talk about is partly determined by the axioms we come to justify, and there is no particular 'absolute' interpretation that we are tending towards (or may miss). This difference, whilst it is unlikely to convince the die-hard Strong Kreiselian Platonist, may be dialectically effective for those of us who remain agnostic on the issue, and also presents a challenge for the Strong Kreiselian to explain how they plan on taking on one of the horns.

We can still accept some implications for the quasi-categoricity theorem even given this picture. For, the quasi-categoricity theorem establishes that *given* an interpretation of the second-order variables, a particular structure is identified by $\mathbf{ZFC}_2$ (with some specific bound on the inaccessibles).[54] We might think that

[54] Multiversists are often explicit on this point. For example Hamkins writes:

> If we make explicit the role of the background set-theoretic context, then the argument appears to reduce to the claim that within any fixed set-theoretic background concept,

this fact has philosophical import. Meadows (2013) identifies three roles for a categoricity theorem:

(1.) to demonstrate that there is a unique structure which corresponds to some mathematical intuition or practice;
(2.) to demonstrate that a theory picks out a unique structure; and
(3.) to classify different types of theory. (Meadows, 2013, p. 526)

He is sceptical about the possibility of (1.) for similar reasons to those I have presented here: The categoricity theorem presupposes the determinateness of the notions it is trying to characterise. However, this is where informal rigour has a role to play; *given* that we have convinced ourselves of informal rigour, the categoricity theorem tells us that our axiomatisation of this notion has been successful. (2.) is thus important; once we believe we have informal rigour, we need to provide a categorical characterisation to manifest this informal rigour (and ensure that the clarity is genuine). I have argued that for set theory, we are not quite there. However, (3.) is important whether or not we actually have informal rigour. The quasi-categoricity theorem for $\mathbf{ZFC}_2$, no matter whether or not we are precise about exactly what structures with boundedly many inaccessibles it concerns, *does* tell us that set theory is non-algebraic. It tells us that our thought at least *aims* at specifying a particular structure, and hence is *not* like concepts and theories of general structure (such as that of *group*) that explicitly aim at dealing with many different non-isomorphic structures. *Inside* every model of $\mathbf{ZFC}_{2QW}$ (which I've argued is possibly the most natural theory for representing our thought about sets), the Zermelo Categoricity Theorem holds and $\mathbf{ZFC}_2$ (with a specific bound on the inaccessibles) is a theory for talking about one isomorphic structure. It is just that this structure can vary across different models of $\mathbf{ZFC}_{2QW}$. Whilst we are not informally rigorous about set theory, the categoricity theorem shows that this situation is *intolerable*, there is *pressure* to become informally rigorous about set theory, even if we currently lack it. This shows that the Difference Thesis (that the case of PP and CH are fundamentally different from one another) can be retained, even in the face of less than full informal rigour in our set concept.

This observation shows that the distinction between *particular* and *general* structures, whilst not incorrect per se, is rather coarse grained. In particular, the idea of *general structure* further subdivides. First, there are those general structures whose concept does not produce a theory for which there is a categoricity proof (e.g. *group*), and thus there is no pressure to hold that informal rigour requires us to determine a particular structure. Call these *intentionally* general structures. There are other concepts (e.g. *set below the first inaccessible*) where we *do* have an axiomatisation with a categoricity proof, even if we *don't* take ourselves to be

any set concept that has all the sets agrees with that background concept; and hence any two of them agree with each other. But such a claim seems far from categoricity, should one entertain the idea that there can be different incompatible set-theoretic backgrounds. (Hamkins, 2012, p. 427)

informally rigorous yet. We call these *unintentionally* general structures. For set theory, whilst we should not take ourselves to have determined a particular structure, there are still *portions* of structures corresponding to this concept that are particular (e.g. the representations of some countable structures within models of theories corresponding to our set concept).

Challenge *How do we know when we reach informal rigour?* In responding to the last objection, I suggested that there are certain concepts (and discourse) about which we are not yet informally rigorous, but there is nonetheless pressure to *become* informally rigorous. This immediately raises the following question: How do we know when we are informally rigorous?

My answer here is a little speculative, but it suggests some interesting directions for future research. We begin with the following idea:

Definition (Informal and Philosophical) We say that a theory **T** exhibits a *high-degree of theoretical completeness* when there are no known sentences other than meta-theoretic sentences (e.g. Gödelian diagonal sentences) independent from **T**.

I acknowledge that this definition is somewhat imprecise. In particular I have no technical account on offer of what is meant by 'meta-theoretic' statements, and I hope that future philosophical research will clarify this notion further. However, it seems that we have *some* handle on the notion though, there seems to be a sense in which $Con(\mathbf{ZFC})$ is a statement of a very different kind from CH.[55]

Given a handle upon the notion, I have the following suggestion; a good indicator[56] of informal rigour is the existence of a categoricity theorem for the relevant second-order theory *and* a high-degree of theoretical completeness (i.e. the only known sentences independent from our theory are obviously meta-theoretic in some way). If this is the case, then if we take ourselves to be informally rigorous and in fact all known independent statements are meta-theoretic, then we can't construct the kind of simplistic argument from the MDA that I've considered here; any known candidate independent statement does not correspond to a legitimate extension of the concept by design.[57]

This is precisely our current situation in arithmetic. Moreover, as mentioned earlier, if we accept Projective Determinacy (which is agreed on by proponents of both Ultimate-L and PFA, since they both think that $\mathsf{AD}^{L(\mathbb{R})}$ holds) then the same situation holds for $H(\omega_1)$; the only sentences about $H(\omega_1)$ that are known

[55] Not least because $Con(\mathbf{ZFC})$ is absolute for well-founded models of **ZFC**, which I've argued our concept of set is sufficient to determine.

[56] I stop short of claiming full sufficiency, simply because I'm not clear that these requirements are sufficient and I don't want to overstate my case. The conjecture that replaces "good indicator" with "sufficient" is still worthy of study.

[57] Walter Dean suggests that this part of my view can be seen as a kind of transcendental refutation of the existence of Orey sentences for a given concept. This seems to be precisely what informal rigour should be aiming at; removing the Orey-phenomenon wherever possible by determining a particular structure.

to be independent from **ZFC** + PD are meta-theoretic in some way.[58] As we've seen, our current set-theoretic concept lacks this feature for questions at the level of third-order arithmetic and above (e.g. CH). It is this that will enable us to avoid examples of the kind given earlier where we consider two different legitimate concept extensions, since our informally rigorous concept should immediately tell us that one or the other extension is illegitimate. Thus, *if* my conjecture that a high-degree of theoretical completeness in combination with a categoricity proof is a good-indication of informal rigour, and *if* we accept PD, and *if* we accept that we do not have a high-degree of theoretical completeness with respect to set theory, *then* this supports the idea that $\mathbf{ZFC}_{2QW}$ is a good axiomatisation of our current level of informal rigour, since those concepts for which we have a high-degree of theoretical completeness can be determined up to isomorphism, and those which do not cannot.

Of course, given the claim that theoretical completeness in combination with categoricity likely yields informal rigour, our belief in informal rigour is *defeasible*. It could be, for example, that we *discover* techniques that allow us to find non-meta-theoretic sentences independent from our current theories of arithmetic and analysis. Hamkins entertains this suggestion:

> 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, 2012, p. 428)

Perhaps then one thinks that my account goes too far: Surely we should not allow arithmetic to fail to be informally rigorous in such a situation? And what of the situation of the Predicative Iterabilist? Doesn't the possibility of their situation show that in fact our discourse involving the reals is not determinate?

[58] See here Woodin (2001) and Welch (2014), for the point about PD implying a high degree of theoretical completeness for $H(\omega_1)$. See Woodin (2017) and Steel (2005) respectively for the point that Ultimate-L and PFA imply PD. Acceptance of PD is somewhat controversial, and not universally agreed upon. Some (e.g. Barton and Friedman 2017, Barton 2020, and Antos et al. 2021) consider versions of the *Inner Model Hypothesis* (IMH), an axiom candidate relying on extensions of the universe that implies that PD is false. An interesting fact, though one that represents a slight digression (and so I don't include it in the main body of the text) is that (i) variants of this axiom can be coded in strong impredicative class theories (see here Antos et al. 2021) without referring to extensions (other than through coding), and (ii) some of these variants imply that there are no inaccessible cardinals in V. A sufficiently strong version of $\mathbf{ZFC}_2$ with one of these axioms added would thus be *fully* (rather than *quasi*) categorical axiomatisation. The IMH unfortunately does not touch CH (and so we could still construct the same MDA-style argument), however there are variations of the IMH (e.g. the *Strong* Inner Model Hypothesis SIMH) that imply that CH fails badly. Despite these complications, there is a large community of set theorists that do regard PD as well-justified (see Koellner 2014 for a summary) and so I set this point aside for the purposes of this paper.

I am quite happy to bite this bullet. If a technique along Hamkins' lines were to be found, I would accept that, after all, our thought concerning arithmetic is not determinate on the basis of the MDA. Given my current evidence however, I find this overwhelmingly unlikely; all such evidence (categoricity, theoretical completeness) seems to indicate that we are informally rigorous, and thus I find it likely that no such technique will be forthcoming.

Even if no such technique concerning the natural numbers is forthcoming, one might object to the idea that our talk concerning the reals is determinate. Recall the case of the Predicative Iterabilist, there we noted that there were different legitimate expansions of arithmetic to form the classical and intuitionistic continuum. Doesn't this (on the basis of the MDA) undermine my claim that our concept of *real number* is indeterminate (and hence $\mathbf{ZFC}_{2QW}$ is undermined)?[59]

I do not find this objection convincing (assuming that we accept Projective Determinacy). Given an utterance of some sentence of the form "the continuum is such that ϕ" we might:

(a) Implicitly have either the classical or intuitionistic (or maybe even infinitesimal) continuum in mind.
(b) Be using "the continuum" as an *algebraic* concept to refer to different non-isomorphic continua.
(c) Have a non-informally rigorous concept of continuum which admits of multiple different inconsistent sharpenings.

If we are using "the continuum" in sense (b), then the objection fails to gain traction, since we are only concerned here with informal rigour as it applies to determination of *particular* structures using *non-algebraic* theories. The fact that there are algebraic uses of the term "continuum" does not affect the fact that I can be precise when talking about a *specific* continuum I have in mind in other contexts when employing a different concept (e.g. the classical continuum).

Further, if we are in case (a) the objection also has no force. If, on a given occasion of utterance, I am clear which specific continuum-concept I am employing (say the classical continuum), then the fact that I can use the word "continuum" to apply to other kinds of continua on different occasions is no more problematic than the fact that I can use the word "pingüino" to refer to a delicious chocolate/cream-based snack as well as a kind of flightless bird.

For the objection to have any force, it must be that we are in case (c). Certainly it is plausible that some agents might find themselves in this position, such as the early analysts or even the average student in beginning a first course in analysis. However as far as contemporary research-level mathematics goes, I think there are some reasons to think that we are not in this position concerning the classical continuum. This is because (as mentioned) earlier, if we accept PD (which is agreed on by both proponents of Ultimate-L and forcing axioms) then are no known sentences for which a MDA-style argument could work. Of course, if our confidence in PD

[59] I thank Geoffrey Hellman for pressing this objection.

were to be challenged (for example by the emergence of a foundational programme rejecting it[60]) then I would be happy to retreat and accept that $\mathbf{ZFC}_{2W}$ should underwrite our axiomatisation of informal rigour (possibly extended to ensure the well-foundedness of the intended structures). In the other direction, it may turn out in fact that there are agents who are *already* informally rigorous using $\mathbf{ZFC}_2$. For example, if we suppose that Ultimate-L comes to be accepted in the next 10 years on the basis of the arguments currently advanced for it, we may wish to conclude that those that currently accept the axiom on these grounds *already* have a theory of sets with a high-degree of theoretical completeness. For now, I remain agnostic regarding current foundational programmes in the philosophy of set theory, and so find MDA-style arguments at least somewhat convincing.

Objection *Mathematics is necessary!* It is very natural at this point to make the following objection: I have claimed that our concept of set is currently not informally rigorous and fails to determine a truth-value for CH. However I've also left open the possibility that in the future we might have an informally rigorous concept of set that determines CH. Moreover, I think that the Axiom of Foundation was not determinate for Mirimanoff's discourse about sets, whereas it is true given our concept of set. But don't I think then that mathematical truth can *vary*? Doesn't this contradict the widely held assumption that mathematical truth is *necessary*? My answer: Yes and no. We can have similar discourses using terms like "set" that are interpreted in very different ways at different times. However once the underlying concept of a discipline is fixed, the truths about that concept at that time are necessary. The only way that truth involving the discourse can vary is by the underlying concepts *changing* somehow.[61] So if by "mathematical truth is necessary" we mean "all truths about every mathematical discourse are fixed" then mathematical truth is not necessary, however if we mean "what is true of particular concepts at particular times is fixed" then mathematical truth *is* necessary.

A comparison case is useful here. Sheldon Smith (2015) argues (convincingly, in my opinion) that Newton's thought involving the concept *derivative* could have been sharpened into several precise non-extensionally-equivalent concepts. Two such are the contemporary conception of *standard* derivative, and the *symmetric* derivative. For the purposes of our discussion it isn't terribly important how these are defined, but they are not extensionally equivalent (for example, if we consider the absolute value function $f(x) = |x|$, the standard derivative is undefined at the origin, whereas it is the constant 0 function (i.e. the x-axis) for the symmetric derivative). Let us suppose (as Smith argues) that Newton's concept $derivative^{\text{Newton}}$ admitted of sharpenings to our concepts *standard derivative* and *symmetric derivative*. Then we should hold that Newton's discourse about the derivative of functions did not

[60] The *Hyperuniverse Programme*, which motivates the Inner Model Hypothesis, is plausibly one such programme.

[61] This idea has much in common with the discussion in Ferreirós (2016) of the idea of *invention cum discovery*.

determine a truth value for the sentence "The derivative of the absolute value function at the origin is the constant 0 function". However, that sentence from our discourse is naturally interpreted (in most contexts) as false, since the concept to be employed (without further specification) for us is *standard derivative*, and the absolute value function has no derivative at the origin for the standard derivative.[62] But we should not think that such an example seriously threatens the idea that mathematical truth is necessary, since the underlying concepts have changed in some way.[63]

Objection *First-order schemas and second-order interpretations.* A key part of Kreisel's 1967 paper is the idea that our commitment to first-order schemas is dependent upon the relevant second-order formulations (e.g. Replacement):

> A moment's reflection shows that the evidence of the first order axiom schema[64] derives from the second order schema: the difference is that when one puts down the first order schema one is supposed to have convinced oneself that the specific formulae used (in particular, the logical operations) are well defined in any structure that one considers... (Kreisel 1967, p. 148.)

His idea is that the informal rigour about the second-order concept is precisely what motivates the first-order schema. Since we are precise about the relevant particular structure, we can see that the first-order schema is always true on this structure, and this is what justifies the principle. Given this claim, and the fact that I have advocated an indeterminacy in the second-order quantifiers in certain contexts, does this undercut the motivation for the first-order schema of Replacement in terms of its second-order formulations?

Kreisel's point is controversial, but even if we accept the idea my response is quick: No. This is because the motivation for the first-order schema could be interpreted as follows: Given *any* particular interpretation of the second-order variables (a notion which here I'm taking to be indeterminate) the first-order schema is true. I do not need to be precise about the interpretation of the second-order variables in order to say that however I interpret them, the instances of the first-order schema hold (this is itself a schematic claim). Kreisel seems to be assuming here that an acceptance of meaningful impredicative second-order theories entails a commitment to determinacy in how the quantifiers are interpreted, but this is a

[62] Thanks here to Zeynep Soysal for suggesting that the concept of derivative might be a pertinent comparison case. See Smith (2015) for the details. That paper also contains several interesting remarks about how we might think conceptual indeterminacy and optimal theories relate in this context, critically examining Rey (1998)'s suggestion that we can implicitly think with a particular concept in virtue of deference to an optimal theory.

[63] Whether or not they are the *same* concept is a question we leave open and will mention in the conclusion.

[64] Here, Kreisel is in fact talking about induction schema in **PA**, but the point transfers to Replacement.

mistake, one can perfectly well accept impredicative second-order theories whilst denying that they have determinate interpretation.[65]

Objection *You've used notions that are dependent upon a definite concept of set in characterising the debate.* A further question is the following: Often I have used phrases like "range of the second-order variables" or "isomorphism" that are naturally interpreted as involving essentially higher-order concepts. But, by my own lights, these notions are indeterminate (for example, I can make two unstructured sets A and B such that $|A| < |B|$ isomorphic by collapsing $|B|$ to $|A|$). How is this legitimate given that I take our talk about sets to be indeterminate?

There are a two points to make here:

First, I *do* take myself to be informally rigorous about a good deal of mathematics (for the purposes of this paper anyway). I think it is likely that we, as a community, are informally rigorous about the real numbers and natural numbers, and the concept of well-foundedness. Thus, my view does not collapse into an 'anything goes' relativism.

Second, we can think of this paper as a modelling exercise concerning what we might be able to say about our current thought in the future. I might begin by saying "Suppose that we were informally rigorous about our concept *set*, what should we then say about our current thought?" I then take myself to have fixed some particular structure $\mathfrak{M}$ about which I am informally rigorous and satisfies $\mathbf{ZFC}_2$ (possibly with a Henkin interpretation!) and analyse how the debate might be interpreted relative to $\mathfrak{M}$ (e.g. that from the perspective of this hypothetical fixed universe,[66] our current thought would be best axiomatised by $\mathbf{ZFC}_{2QW}$). This will then resemble how our intellectual descendants who are informally rigorous (should there be any) might think of our thought, much as how we now look at Mirimanoff's thought as indeterminate.

6.7 Conclusions and Open Questions

In this paper, I've argued that there are various foundational programmes and situations we might find ourselves in that support different levels of informal rigour concerning our set-theoretic concepts and thought. In particular, I've suggested that our level of informal rigour in set theory might be insufficient to convince us that our discourse and concepts determine a particular set-theoretic structure. Instead, perhaps we should admit some structural relativity into our characterisations of structures, and a logic weaker than second-order is appropriate for characterising

[65] This point has been made increasingly vivid by the recent boom in the study of different *class*-theoretic systems.

[66] There are options here for how we might interpret this reference. It might be interpreted as picking out a specific such $\mathfrak{M}$ (as outlined in Breckenridge and Magidor 2012) or an 'arbitrary' such $\mathfrak{M}$ in the style of Fine and Tennant (1983). See Horsten (2019) for a recent treatment.

our current thought about sets (in particular $\mathbf{ZFC}_{2QW}$). I've also argued, however, that there is pressure on us to develop a more informally rigorous concept of set, and thereby answer questions like CH. This identifies a fundamental distinction among the general structures; we have structures that are *unintentionally* general (like the structure corresponding to our discourse about sets) and those that are *intentionally* general (like the group structure). This said, there are lots of questions left open by the paper. I take this opportunity to raise some of the main ones.

Question. What is the status of the Modal Definiteness Assumption?

For most of the paper, I was happy to take the MDA as an assumption. I think that given the kinds of possibilities described in the paper (Mirimanoff's futures, and our own) it's a very plausible assumption. This said, I am pretty convinced that both Kreisel and Isaacson would be unhappy with it (since it obviously implies their position concerning the determinateness of CH is false), and I haven't subjected it to really intense philosophical scrutiny. This is worth examination.

A second question concerns the kinds of particular structures determined by our set-theoretic discourse and concepts. Assuming that I am right that $\mathbf{ZFC}_{2QW}$ is the right axiomatisation of our current discourse concerning set theory, there is the question of what is determined on this basis. By and large this theory has not (to my knowledge) been studied in detail.[67] There are some clear candidates for particular structures that can be given categorical characterisations given an acceptance of $\mathbf{ZFC}_{2QW}$, we have already mentioned $H(\omega_1)$ and $(\mathbb{R}, +, \times, <)$. However, there are others; the Shepherdson-Cohen minimal model for example can be given a categorical axiomatisation, since we can capture absolutely the notion of well-foundedness. The theory consisting of the following axioms:

(i) **ZFC**-Foundation
(ii) The Axiom of Foundation formulated as the sentence (in quasi-weak second-order logic) that there are no infinite descending $\in$-chains.
(iii) $V = L$
(iv) $\neg\exists\mathfrak{M}$"$\mathfrak{M}$ is a transitive model of **ZFC**"

identifies a unique model up to isomorphism, since the Shepherdson-Cohen minimal model is (assuming that there is a transitive model of **ZFC**) the unique transitive model of **ZFC** satisfying $V = L$ and containing no transitive models of **ZFC**. An anonymous reviewer helpfully points out that for other countable structures what can be determined up to isomorphism may depend on ambient facts about independence. For example, if we allow non-recursive axiomatisations, a result of Victor Marek states that if there is a projective well-ordering of the reals (e.g. under $V = L$) then every countable structure is categorical in second-order logic, and hence also categorical in $\mathbf{ZFC}_{2QW}$ (since on countable structures quasi-weak

[67] Much of what I've considered here was gleaned from Shapiro (1991, 2001). A recent contribution that briefly considers some other versions of **ZFC** with different underlying logics is Kennedy et al. (2021) (esp. §8: Semantic Extensions of **ZFC**).

second-order logic coincides with full second-order logic). However, this result is independent of **ZFC**; it is consistent with **ZFC** that there are countable ordinals whose second-order theory is not categorical.[68] We therefore ask:

> **Question.** What other structures (both set-theoretic and non-set-theoretic) are particular, *given* that we accept that our thought is axiomatised by $\mathbf{ZFC}_{2QW}$, and how can we provide a bottom-up characterisation for them?

Closely related is whether or not the only unintentionally general structures we talk about are set-theoretic. For all I've said, it might just be set theory that exhibits this feature. We might then ask:

> **Question.** Are there other interesting unintentionally general structures apart from set-theoretic ones?

Throughout the paper, I talked of concepts changing, for example in the shift from Mirimanoff's concept to our own, from Newton's concept of derivative, and from our own concept of set to that of our intellectual descendants. An interesting philosophical question is then in what sense there is a *continuity* of conceptual content and thought between one intellectual generation and the next. We therefore ask:

> **Question.** When a concept is made more precise, what remains constant, and how should we understand this continuity? Does the concept *change* or should we rather understand this as a shift to a *different* concept? Given this, in what sense do we *mean the same/similar thing(s)* by what we say with our mathematical utterances?[69]

We save the toughest question for last. Throughout, I've talked as though we might one day be informally rigorous about our concept of set. However, this might just not be possible. Perhaps any modification of the concept we suggest will be susceptible to decisive objections. Perhaps the different possibilities for extending our concept of set will all seem equally legitimate, and we simply cannot reasonably pick any one concept, whatever the pressure from the quasi-categoricity theorem.[70] We therefore ask:

> **Question.** Is it possible for us to legitimately develop an informally rigorous concept of set (at least for each level of the hierarchy)?

---

[68] I am very grateful to an anonymous reviewer for explaining these facts and pointing me to the discussion on Mathoverflow at Sáez (2011) and Schweber (2014), as well as the mentioned result in Marek (1973).

[69] I am grateful to Chris Scambler for proposing this question and some interesting discussion here. Some possible directions of research (suggested to me by Fenner Tanswell and Juliette Kennedy) include revisiting Lakatos (1976), and in particular development this idea using Waismann's notion of *open texture* and resources from *conceptual engineering*. This has been examined in the case of the Church-Turing thesis by Shapiro (2013), but also in the philosophy of mathematics more broadly in Tanswell (2018) and by Vecht (Forthcoming), with the former providing an application to the universe/multiverse debate.

[70] Considerations along these lines are considered in Hamkins (2012, 2015).

Perhaps we can answer this question affirmatively, or perhaps we are doomed to spend our days like a mathematical version of Buridan's Ass, trapped between equally (un)attractive options. Time will tell.

Acknowledgments I would like to thank Carolin Antos, Andy Arana, John Baldwin, Hans Briegel, Mirna Džamonja, Walter Dean, Monroe Eskew, Ben Fairbairn, José Ferreirós, Sarah Hart, Geoffrey Hellman, Deborah Kant, Juliette Kennedy, Daniel Kuby, Hannes Leitgeb, Julia Millhouse, Moritz Müller, Thomas Müller, Gianluigi Oliveri, Georg Schiemer, Daniela Schuster, Fenner Tanswell, Zeynep Soysal, Jouko Väänänen, Giorgio Venturi, Matteo Viale, Andrés Villaveces, Verena Wagner, John Wigglesworth and audiences in Konstanz, Munich, Paris, and Vienna for helpful discussion. Special mention must be made of Chris Scambler—many ideas in the paper arose out of work on a joint project with him, and I am grateful to him for permission to include them (mistakes made in filling out the details are my own). Two anonymous reviewers provided comments that greatly helped improve the paper, and I am very grateful for their close reading and useful remarks. I am also very grateful for the generous support of the FWF (Austrian Science Fund) through Project P 28420 (*The Hyperuniverse Programme*) and the VolkswagenStiftung project *Forcing: Conceptual Change in the Foundations of Mathematics*.

References

Aczel, P. 1988. *Non-well-Founded Sets*. Stanford: CSLI Publications.

Antos, C., N. Barton, and S. Friedman. 2021. Universism and extensions of V . *The Review of Symbolic Logic* 14(1): 112–154. https://doi.org/10.1017/S1755020320000271.

Barton, N. 2020. Forcing and the universe of sets: Must we lose insight? *Journal of Philosophical Logic* 49: 575–612.

Barton, N., and S.-D. Friedman. 2017. Maximality and ontology: How axiom content varies across philosophical frameworks. *Synthese* 197(2): 623–649.

Ben-David, S., P. Hrubeš, S. Moran, A. Shpilka, and A. Yehudayoff. 2019. Learnability can be undecidable. *Nature Machine Intelligence* 1(1): 44–48.

Boolos, G. 1971. The iterative conception of set. *The Journal of Philosophy* 68(8): 215–231.

Breckenridge, W., and O. Magidor. 2012. Arbitrary reference. *Philosophical Studies* 158(3): 377–400.

Button, T., and S. Walsh 2018. *Philosophy and Model Theory*. Oxford University Press.

Corry, L. 2004. *Modern Algebra and the Rise of Mathematical Structures*. Birkhäuser.

Dedekind, R. 1888. Was sind und was sollen die Zahlen? In Ewald, W.B., ed. 1996a. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, vol. II, 787–832. Oxford University Press.

Dummett, M. 1977. *Elements of Intuitionism*. Oxford University Press.

Feferman, S., and G. Hellman. 1995. Predicative foundations of arithmetic. *Journal of Philosophical Logic* 24(1): 1–17.

Ferreirós, J. 2016. *Mathematical Knowledge and the Interplay of Practices*. Princeton University Press.

Fine, K., and N. Tennant. 1983. A defence of arbitrary objects. *Proceedings of the Aristotelian Society* 57: 55–77, 79–89.

Gödel, K. 1940. *The Consistency of The Continuum Hypothesis*. Princeton University Press.

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

Hamkins, J.D. 2015. Is the dream solution of the continuum hypothesis attainable? *Notre Dame Journal of Formal Logic* 56(1): 135–145.

Hekking, J. 2015. Natural models, second-order logic & categoricity in set theory. Bachelor's Thesis, Leiden University.

Hellman, G. 1996. Structuralism without structures. *Philosophia Mathematica* 4(2): 100–123.
Hellman, G., and S. Feferman. 2000. *Challenges to Predicative Foundations of Arithmetic*, 317–338. Cambridge University Press.
Horsten, L. 2019. *The Metaphysics and Mathematics of Arbitrary Objects*. Cambridge University Press.
Incurvati, L. 2014. The graph conception of set. *Journal of Philosophical Logic* 43(1): 181–208.
Isaacson, D. 2011. The reality of mathematics and the case of set theory. In *Truth, Reference, and Realism*, ed. Z. Noviak and A. Simonyi, 1–75. Central European University Press.
Kanamori, A. 1996. The mathematical development of set theory from Cantor to Cohen. *The Bulletin of Symbolic Logic* 2(1): 1–71.
Kanamori, A. 2004. Zermelo and set theory. *Bulletin of Symbolic Logic* 10(4): 487–553.
Kennedy, J., M. Magidor, and J. Väänänen. 2021. Inner models from extended logics: Part 1. Journal of Mathematical Logic 21(2).
Koellner, P. 2014. Large cardinals and determinacy. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta, Spring 2014 ed. Metaphysics Research Lab, Stanford University.
Kreisel, G. 1967. Informal rigour and completeness proofs. In *Problems in the Philosophy of Mathematics*, ed. I. Lakatos, 138–186.
Lakatos, I. 1976. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press.
Leitgeb, H. 2020. On non-eliminative structuralism. Unlabeled graphs as a case study (Part A). *Philosophia Mathematica* 28: 317–346
Marek, W. 1973. Consistance d'une hypothèse de fraïssé sur la définissabilité dans un language du second ordre. *C. R. Acad. Sci. Paris Sér. A-B* A–B276: A1147–A1150.
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.
Mirimanoff, D. 1917. Les antinomies de Russell et de Burali-Forti et le probleme fondamental de la theorie des ensembles. *L'Enseignement Mathématique* 19: 37–52.
Reck, E., and G. Schiemer. 2019. Structuralism in the philosophy of mathematics. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta, Winter 2019 ed. Metaphysics Research Lab, Stanford University.
Resnik, M. 1997. *Mathematics as a Science of Patterns*. Oxford University Press.
Rey, G. 1998. What implicit conceptions are unlikely to do. *Philosophical Issues* 9: 93–104.
Rumfitt, I. 2015. *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford University Press.
Sáez, C. 2011. Categoricity in second order logic. MathOverflow. https://mathoverflow.net/q/72635 (version: 2011-08-10).
Scambler, C. Categoricity and determinacy. Manuscript under review.
Schweber, N. 2014. The (non-)absoluteness of second-order elementary equivalence. MathOverflow. https://mathoverflow.net/q/161676 (version: 2014-03-28).
Shapiro, S. 1991. *Foundations without Foundationalism: A Case for Second-order Logic*. Oxford University Press.
Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press.
Shapiro, S. 2001. *Systems Between First-Order and Second-Order Logics*, 131–187. Dordrecht: Springer.
Shapiro, S. 2013. The open texture of computability. In *Computability: Turing, Gödel, Church, and Beyond*, ed. B.J. Copeland, C.J. Posy, and O. Shagrir, 153–181. Cambridge, MA: MIT Press.
Shepherdson, J.C. 1951. Inner models for set theory–Part I. *Journal of Symbolic Logic* 16(3): 161–190.
Shepherdson, J.C. 1952. Inner models for set theory–Part II. *Journal of Symbolic Logic* 17(4): 225–237.

Shepherdson, J. 1953. Inner models for set theory–Part III. *The Journal of Symbolic Logic* 18(2): 145–167.

Smith, S.R. 2015. Incomplete understanding of concepts: The case of the derivative. *Mind* 124(496): 1163–1199.

Steel, J.R. 2005. PFA implies $\mathsf{AD}^{L(\mathbb{R})}$. *The Journal of Symbolic Logic* 70(4): 1255–1296.

Tanswell, F.S. 2018. Conceptual engineering for mathematical concepts. *Inquiry* 61(8): 881–913.

Taylor, W. 2019. Learnability can be independent of ZFC axioms: Explanations and implications.

Väänänen, J., and T. Wang. 2015. Internal categoricity in arithmetic and set theory. *Notre Dame J. Formal Logic* 56(1): 121–134.

Vecht, J. Forthcoming. Open texture clarified. *Inquiry*. https://doi.org/10.1080/0020174X.2020.1787222.

Welch, P. 2014. Global reflection principles. Isaac Newton Institute pre-print series, No. NI12051-SAS.

Woodin, H. 2001. The continuum hypothesis, Part I. *Notices of the American Mathematical Society* 48(6): 569–576.

Woodin, W.H. 2017. In search of Ultimate-*L*: The 19th Midrasha Mathematicae lectures. *The Bulletin of Symbolic Logic* 23(1): 1–109.

Zermelo, E. 1930. On boundary numbers and domains of sets. In Ewald, W.B., ed. 1996b. From Kant to Hilbert. *A Source Book in the Foundations of Mathematics*, vol. I, vol. 2, 1208–1233. Oxford University Press.

Chapter 7
Ontological Dependence and Grounding for a Weak Mathematical Structuralism

Silvia Bianchi

Abstract In the philosophy of science, Weak Structural Realism (WSR) offers a promising priority-based strategy to avoid the main objection to eliminative Ontic Structural Realism (OSR). On that view, quantum particles *depend* for their identity on quantum entanglement structures but are defined as not entirely structural *thin physical objects*. A similar approach can be applied to mathematical structuralism, where Weak Mathematical Structuralism (WMS) provides a novel, more moderate interpretation of *ante rem* structuralism. WMS is articulated in terms of grounding: numbers are *grounded* for their identity in the abstract structure they belong to. However, they are not completely reduced to their structural features and are re-conceptualized as *thin mathematical objects*, endowed with both structural and non-structural properties. The introduction of such objects in the structural ontology allows to escape some typical objections to *ante rem* structuralism without abandoning the priority of structures.

Keywords Scientific structuralism · Mathematical structuralism · Ontological dependence · Metaphysical grounding · Thin objects · Individuation

7.1 Introduction

Ontic Structural Realism aims at providing the best interpretation of scientific realism, according to which 'there are no things and structure is all there is' (Ladyman and Ross 2007) – or, at least, all there is *fundamentally*. This view has interesting connections with Shapiro's *ante rem* structuralism in the philosophy of mathematics, which assumes a background ontology of abstract structures and reduces mathematical objects to mere positions/empty places in these structures.

S. Bianchi (✉)
University School for Advanced Studies (IUSS), Pavia, Italy
e-mail: silvia.bianchi@iusspavia.it

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_7

Both accounts share an entirely structural conception of objects, which raises related issues in scientific and mathematical structuralism: scientific OSR is subject to 'the relation without *relata*' objection, dealing with the nature of quantum particles in quantum entanglement structures. Mathematical *ante rem* structuralism meets with the 'problem of objects' and the 'problem of identity' concerning the individuation of numbers in abstract structures.

The core idea is to resist these objections by reconsidering the relationship between objects and structures in terms of *ontological dependence* and *metaphysical grounding* (objects are dependent on/grounded in the relevant structures); these notions display some features which naturally support a non-eliminative stance towards objects, in which they are given a more substantial role. In the philosophy of science, weaker forms of structuralism (Esfeld 2004; Wolff 2012) have been already introduced as alternatives to OSR. I will specifically refer to their formulation in terms of Lowe's (1994, 2016) identity dependence, which allows to introduce *thin physical objects* in the structural ontology. Such objects – albeit secondary to structures – are not entirely reducible to their structural features and suggest a possible response to the *relation without relata* objection affecting OSR.

My main purpose is to consider the theoretical advantages of a weak approach also in the mathematical framework, where it has not been explicitly proposed. In analogy with the philosophy of science, I will elaborate Weak Mathematical Structuralism (WMS) as a more moderate interpretation of Shapiro's non-eliminative *ante rem* structuralism. WMS is considerably based on Linnebo's (2008) and Wigglesworth's (2018) proposals, which appeal to the notions of identity dependence and metaphysical grounding. The stricter connection between grounding and metaphysical explanation makes grounding particularly suitable to express non-eliminative structuralism and, specifically, WMS. Moreover, grounding fits well with the prospects of applying a non-eliminative approach to both objects and structures; on the one hand, numbers are defined as not-entirely structural *thin mathematical objects*, comparable with thin physical objects. On that view, a possible response to the main difficulties of Shapiro's *ante rem* structuralism is put forward. On the other hand, the priority of structures is preserved, consistently with *ante rem* structuralism.

The present paper is structured as follows: the first part deals with scientific structuralism; first of all, I will present Ontic Structural Realism and its main objection (Sect. 7.2). In Sect. 7.3, I will illustrate French's taxonomy of OSR-positions in terms of dependence, showing how this notion favors Weak Structural Realism (WSR) as opposed to eliminative OSR and Moderate Structural Realism (MSR). On this basis, a more detailed characterization of WSR and thin physical objects will be advanced as a possible way of escaping OSR's objection (Sect. 7.4).

The second part of the discussion concerns mathematical structuralism. In Sect. 7.5, Shapiro's *ante rem* structuralism and its main difficulties will be illustrated; I will then take into account the formulations of Shapiro's account in terms of dependence (Linnebo 2008) and grounding (Wigglesworth 2018) (Sect. 7.6). By focusing on grounding, I will develop my own account of Weak Mathematical Structuralism (WMS) and thin mathematical objects, which avoid the 'problem of

identity' and the 'problem of objects' of *ante rem* structuralism without abandoning an *ante rem* individuation of structures (Sect. 7.7).

7.2 Ontic Structural Realism (OSR) and the 'Relation Without *relata* Objection'

Ontic Structural Realism (OSR) claims to offer the best metaphysical interpretation of our contemporary physics (concerned with quantum particles and the quantum entanglement structures they are in) and comprises a family of views; on its broadest interpretation, it is committed to the *fundamentality* of structures and their *priority* to objects. This means that «the fundamental ontology of the world is one of structures and that objects, as commonly conceived, are at best derivative, as worst eliminable». (French 2014, p. v). Different attempts to make the priority and the fundamentality claims more precise have been proposed. Among them, *Eliminative OSR* (French and Ladyman 2003; Ladyman and Ross 2007; French 2010) replaces the object-oriented metaphysics of the received view – which is not vindicated by our present understanding of Quantum Mechanics (QM) – with a picture in which objects are eliminated *tout court*. In slogan form, «there are no things, and structure is all there is» (Ladyman and Ross 2007, p. 131). This metaphysical shift is motivated by a fundamental underdetermination concerning the individuality of quantum particles in QM – consistent with two alternatives metaphysical packages: quantum particles as individuals and as not individuals.[1] The original OSR's contributions are intended to break such underdetermination by introducing a third way, in which the concept of object itself is undermined and quantum particles are reconceptualized in purely structural terms. Metaphysically speaking, this yields the result that all that matters about quantum particles are their *structural properties*.

Eliminative OSR appears seriously controversial and subject to the *relation without relata objection* (Cao 2003; Dorato 2000; Psillos 2001, 2006; Busch 2003; Morganti 2004; Chakravartty 1998, 2003), which questions how we can have a structure without the individuals making up this structure. In Chackravartty's (1998, p. 399) words «one cannot intelligibly subscribe to the reality of relations unless one is also committed to the fact that some things are related». Eliminativists endeavoured in making sense of the 'relations without *relata*' intuition by interpreting structures as universals (Stein 1989; Psillos 2006) or arguing that the *relata* of the relations turn out to be structures themselves (Ladyman and Ross 2007; Saunders 2003).

Either way, such proposals are largely contentious, leaving room to the more defensible non-eliminative OSR, which includes objects in the structural ontology. The non-eliminative approach comes into a variety of forms; some of them

[1] Cf. French and Krause (2006).

preserve the priority of structures and understand objects as secondary to them (priority-based strategies) whereas others posit objects and structures on the same fundamental level (parity-based strategies). Along the lines of French (2010), I will focus on Moderate Structural Realism (MSR) and Weak Structural Realism (WSR) as the most full-fledged examples, which can be accounted for in terms of ontological dependence.

7.3 OSR and Dependence

In scientific structuralism, a more fine-grained analysis of the relation between objects and structures remains to be accomplished. According to French (2010, p. 98) «there is a lack of clarity regarding the relationship between objects and structures, and it is also one that effects a separation between the eliminativist and non-eliminativist forms of ontic structural realism».

Ontological dependence seems to adequately fill this gap, by cashing out the priority and the fundamentality claims at hand in OSR. Broadly speaking, ontological dependence is a metaphysical and explicative notion which conveys a distinctively non-causal priority relation among entities; in particular, an entity is said to be dependent on another entity either for its *existence* or for its *identity*.

Significantly, different forms of dependence elucidate different forms of OSR, providing a taxonomy in which eliminative (OSR), Moderate Structural Realism (MSR) and Weak Structural Realism (WSR) are distinguished.[2]

First, French (2010, p. 106) outlines eliminative OSR as follows:

1. OSR: the very constitution (or essence) of the putative objects is dependent on the relations of the structures.

This intuition is clarified by Fine's (1995) essential-existential account of dependence (EDE):

x depends $_{\mathrm{E}}$ for its existence upon $y =_{\mathrm{df}}$ it is part of the essence of x that x exists only if y exists.

In eliminative OSR objects are radically reduced to their structural features: they solely exist if the relevant structure exists and there is nothing to them (identity, constitution, etc.) which can be defined independently of the structure. After all, objects are not genuine *relata* of the dependence relation under scrutiny and it is questionable whether a dependence relation applies at all.

Second, Moderate Structural Realism (MSR) introduces a mutual relation of dependence between objects and structures, that are ontologically on a par. As explained by French (2010, p. 104):

[2] French specifically refers to Tahko and Lowe's (2016) analysis of dependence, in which different accounts are illustrated.

2. MSR: the identity of the objects/nodes is (symmetrically) dependent on that of the relations of the structure and *viceversa*.

This conception is expressed by the modal-existential account of dependence (EDR), which allows for symmetrical relations[3] and is laid out by Lowe (2005) as follows:

x depends $_{R}$ for its existence upon $y =_{df}$ necessarily, x exists only if y exists.

Moderate Structural Realism (Esfeld and Lam 2008) states that both objects and structures are ontologically fundamental entities – the ontological priority of structures is reformulated in terms of a 'parity claim', where objects and structures are on the same ontological footing.[4]

> Neither objects nor relations (structure) have an ontological priority with respect to the physical world: they are both on the same footing, belonging both to the ontological ground floor. (Esfeld and Lam 2008, p. 31).

Third, WSR understands the relation between objects and structures *asymmetrically* (objects are admitted in the ontology, though as less fundamental than the structures they belong to) and focuses on the notion of identity. French describes WSR as follows (p. 105):

3. The identity of the putative objects/nodes is (asymmetrically) dependent on that of the relations of the structure.

According to French, the asymmetrical notion of dependence at play is adequately captured by Lowe's (1994, 2016) identity dependence (ID):

x depends for its identity upon $y =_{df}$ there is a two-place predicate "F" such that it is part of the essence of x that x is related by F to y.

In scientific structuralism, WSR preserves the priority of structures and is related to the idea of a *contextual identity* for quantum particles, derivative on the relations in which they stand (Stachel 2002; Ladyman 2007) and sufficient to support a *thin* notion of objects. The idea of thin objects has been originally elaborated by Saunders (2003) in terms of a weaker form of the Principle of Identity of Indiscernibles (PII) and a weak notion of discernibility for quantum particles. Such proposal is based on Quine's (1960, p. 230) distinction between different grades of discernibility, consisting of *absolute*, *relative* and *weak discernibility*. Two objects are absolutely discernible if there is a one-variable formula which is true of an object and not of another; relatively discernible if there is a two free-variables formula which applies to them just in one order; weakly discernible if there is a symmetrical but irreflexive relation holding between them. Let us now come back

[3] Tahko and Lowe (2016, sec. 2.1.) consider as an example the relation between Socrates and its life, which are said to be dependent on each other.

[4] Such view is generally supported by a relational interpretation of properties and by a discussion of quantum entanglement in terms of non-separability.

to quantum particles' relevant case, and consider two electrons in a singlet-state having an opposite spin: while the two particles cannot be either absolutely or relatively discernible (they are indistinguishable in isolation, since their permutation leaves the state they are in unchanged) they are *weakly discernible* in virtue of the irreflexive relation holding between them (i.e. having opposite direction of each component of spin to . . .). This proposal turns out to be partially controversial in the structuralist literature, since structures, in order to individuate the *relata*, seem to presuppose their numerical diversity, and then cannot account for it.[5] However, other interpretations of thin objects, focused on a more accurate analysis of the notion of dependence at play, are available to WSR (Sect. 7.4.)

Before moving on, a reflection on ontological dependence itself allows to evaluate the tenability of each form of OSR. Wolff (2012, p. 608) discusses the role of ontological dependence in scientific structuralism and argues that «only certain forms of structural realism can be articulated using ontological dependence». Significantly, ontological dependence cannot serve the purpose of accounting for an eliminative interpretation of the relation between objects and structures. Ontological dependence, in fact, is not an eliminative relation and requires that both the *relata* of the relation (objects and structures) should exist. To say that B ontologically depends on A means that A is *prior* to B, which is less fundamental than A; but this does not mean that B is to be eliminated . By these means, ontological dependence provides further reasons to reject eliminative OSR (1) and favors non-eliminative views, thus leaving us with options (2) and (3).

Moderate Structural Realism (2) includes objects in the ontology, but is open to criticisms with respect to the symmetry of the notion of dependence at hand. Such assumption is in contrast with the standardly 'layered' metaphysical picture of reality (in which entities come into different levels of fundamentality) and is subject to the typical objections concerning circular explanations.[6]

Hence, Weak Structural Realism (3), as captured by Lowe's (asymmetrical) identity dependence (ID), appears to be the most compelling, priority-based alternative to eliminative OSR:

> Of the three versions of ontic structural realism discussed at the beginning [the three forms recognized by French 2010, *Ed.*] only thin-object OSR comes close to being articulated using essential dependence as the relation between objects and structure» (Wolff 2012, p. 622).

In the next section, I will more specifically define this account and propose a different interpretation of the underlying conception of thin objects, which suggests a plausible response to OSR's *relation without relata* objection.

[5] Analogous considerations apply to the mathematical framework and meet with similar difficulties (cf. sec. 5).

[6] Lowe (2012) raises a more specific objection, according to which a coherent structuralist ontology should include at least some self-individuating entities.

7.4 Weak Structural Realism (WSR) and Quantum Particles as Thin Physical Objects

From the previous discussion, WSR has emerged as the most defensible form of OSR. However, a more precise definition of thin objects, as presupposed by WSR, stands in need of further clarification.

As above, Saunders's (2003) defines thin objects by referring to the Quinean distinction among different grades of discernibility; however, such characterization proves to be troublesome in the scientific domain. My intention is to explore an alternative conception of thin objects, which is arguably immune from the objections to Saunders' (2002) proposal.

To do so, let us rehearse the core assumption of WSR:

1. The identity of the putative objects/nodes is (asymmetrically) dependent (ID) on that of the relations of the structure.

Providing that the notion of dependence at hand is understood in a non-eliminative way, consistently with considerations put forward in Sect. 7.3, this claim *per se* motivates the introduction of thin objects in the ontology, without necessarily appealing to symmetrical and irreflexive relations holding between them.

Esfeld (2004) and Wolff (2012) provide further reasons to articulate a notion of thin objects in these terms, relying on ontological dependence and a more precise investigation of the structural and non-structural properties of quantum particles. Esfeld (2004, p. 613) argues for a non-eliminative metaphysics of relations for quantum particles as follows:

> relations require things that stand in the relations (although these things do not have to be individuals and they need not have intrinsic properties).

This idea allows interpreting physical theories as referring to *entities* that may exist independently of the relations in which they stand. To be an *entity* is to be the subject of a predication of properties. This is not equivalent to be an *individual*, for which further requirements need to be fulfilled (having an intrinsic identity or a 'primitive thisness'). This distinction echoes the traditional opposition between entities on the one hand, and (individual) objects on the other hand, which qualify as 'properly individuated entities'.[7] On this view, is not implausible to consider entities which are not individuals. In WSR, quantum particles in entanglement states are clearly not individual objects in a proper sense, since their identity is entirely determined by the whole entanglement system they are in – as the relation of dependence (ID) at play shows. However, as pointed out by Esfeld (2004), relations presuppose objects of some sort, i.e. *not individuated entities* or, more specifically, *thin objects*.

If the identity of thin objects is given by the structure, what does their existence – conceivable independently of the structure – exactly amount to? I submit the

[7] Cf. Keränen (2001, p. 313)

existence of thin objects as not reduced to their essential structural properties, since it also results in their *non-structural properties*. In scientific structuralism, structural properties of quantum particles (state-dependent properties such as position and momentum) are generally described as those properties which remain invariant under symmetry groups transformations in group theory.[8]

As specified by Ladyman (2020, sec. 4.1.):

> We have various representations of some physical structure which may be transformed or translated into one another, and then we have an invariant state under such transformations which represents the objective state of affairs.

However, this definition admits counter-examples, consisting of the state-independent/non-structural properties of quantum particles. Wolff (2012, p. 623) refers to *kind properties* (such as charge, spin, mass), which qualify particles as electrons, muons, etc. and are not easily reducible to a structural, group-theoretic interpretation:[9]

> Particles *qua* individuals are thin objects. To the extent that we understand their identity as individuals, we understand it in terms of the state they are in. This leaves unaffected their 'kind identity', that is, their identity as electrons rather than muons. Which kind of particles they are does not depend on any particular state the particles are in.

Some clarifications concerning kind properties are needed, since they play a crucial role in the present discussion. First, non-structural kind properties do not correspond to intrinsic properties, defined as the «properties that are independent of whether the object is alone or accompanied by other objects» (Esfeld and Lam 2011, p. 144):[10] in fact, kind properties do not fix the identity of objects as individuals, but as 'packaged' into kinds (given by determinate correlations of mass, spin, charge). Second, properties such as mass and spin are essential properties of quantum particles. This may constitute a possible objection to WSR (and, plausibly, to any form of OSR, in which essential properties are generally understood as structural properties). However, there is a sense in which kind properties are *secondary* to structural properties. Kind properties distinguish electrons, muons, etc. but, assuming two electrons in a singlet-state, they leave underdetermined which one is which. It follows that quantum particles are indistinguishable in a much stronger sense, and that solely the structure fixes their very identity – what they are as opposed to all the other objects in the same structure.

[8] Symmetry groups transformations are specifically presupposed by Quantum Field Theory (QFT) and represent the mathematical counter-part of quantum entanglement states. This interpretation traces back to Cassirer, Born, Weyl and Eddington.

[9] Several attempts to apply this reduction have been performed (cf. Wigner's, 1939, original characterization of properties; more recently, Castellani 1998). However, these structuralist strategies proved to be largely unsuccessful.

[10] Still, this is controversial; for a discussion about kind properties and intrinsic properties see McKenzie (forthcoming).

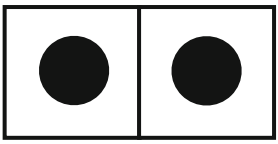


Fig. 7.1 Two quantum particles in a singlet-state

On this basis, let us define thin physical objects by referring to the conjunction of the following conditions, focusing on the *identity* and the *existence* of quantum particles respectively.

1.a) Thin objects [Identity]: thin objects are entities whose *identity* entirely depends upon the relevant structure.
2.a) Thin objects [Existence]: thin objects are entities whose *existence* (resulting in both structural and non-structural kind properties) is necessary to posit relations themselves.

The permutation of quantum particles in a single state provide us with a more specific example, which sheds light on definitions 1.a and 2.a (Fig. 7.1):

Quantum particles are indistinguishable in isolation: they can be permuted while leaving the relevant state unchanged. Therefore, solely quantum entanglement structure grounds their identity as individuals (definition 1.a). Nevertheless, relations of quantum entanglement require *things* to stand in the relations: as a consequence, the relevant particles cannot collapse in a single one, because they also possess state-independent kind properties that allow them to be considered numerically distinguished *relata*; in fact, even though non-structural properties cannot distinguish particles of the same kind, they are able to distinguish particles that belong to different kinds, e.g. electrons, muons, etc. (definition 2.a).

Still, thin objects so understood raise two main worries:

(i) are thin physical objects substantial enough to avoid resulting in a 'no-objects-at-all' position?
(ii) are thin physical objects weak enough to preserve a structuralist framework?

First, as opposed to eliminative OSR, a (weak) notion of object is re-established: thin objects are not entirely reduced to their structural features, since they are endowed with both structural and non-structural *kind properties.* Such properties define them – if not as individuals – as entities which are conceivable independently of the structures and exist metaphysically prior to them.

On these grounds, thin objects are *substantial enough* to be admitted as legitimate *relata* of the structural relations, in accordance with the idea that relations need some things to be related (the first *i*. condition is satisfied). This suggests a plausible response to the 'relation without *relata*' objection affecting OSR.

Second, consistently with OSR, quantum particles do not possess intrinsic properties, which would commit to the object-oriented metaphysics that scientific structuralism wants to contrast. In fact, it is widely held in QM formalism that quantum particles – when entangled – lack any quantum pure state, which is

exhibited just by the whole entangled system. In such cases, quantum particles are devoid of any properties that may characterize them individually, or of any «properties that are independent of whether the object is alone or accompanied by other objects» (Esfeld and Lam 2011). However, this is not always the case; at least on some interpretations of QM (i.e Copenhagen-type interpretations) quantum particles are not necessarily in entanglement states, and then display a pure state in isolation. So, it is worth emphasizing that the 'thinnes' of quantum particles entails a specific two-place relation, involving a particle and a time. Taking into account these two possibilities (quantum particles as entangled/non-entangled), I focus on the more standard situation described by WSR, in which quantum particles are in fact entangled and then depend on the structure for their identity. This makes quantum particles secondary to the structures (thin objects are *weak enough,* thus responding to *ii.*) and reinforces the priority of structures, as required by the asymmetry of WSR.

To sum up, such objects appear to be something more than the 'no-object at all' (whatever *thick* or *thin*) the eliminative versions of OSR are committed to, but something less than the 'thicker' objects in opposition of which OSR has been originally introduced.

Let us move to mathematical structuralism and the analysis of *ante rem* structuralism, where very similar issues arise.

7.5 Shapiro's *ante rem* Structuralism: The 'Problem of Objects' and the 'Problem of Identity'

The theoretical core of mathematical structuralism can be adequately introduced by considering the following quotation from Hellman and Shapiro (2019, p.1):

> The theme of structuralism is that what matters to a mathematical theory is not the internal nature of its objects – numbers, functions, functionals, points, regions, sets, etc. – but how these objects relate to each other.

In what follows, I will focus on the interpretation of this assumption according to Shapiro's (1997) *ante rem* structuralism, which is strictly connected with the discussion of OSR in the philosophy of science.

Shapiro's *ante rem* structuralism aims at introducing a structuralist position which combines realism in ontology (mathematical entities exist) and realism in semantics (mathematical statements have not-vacuous truth values) with an acceptable epistemology, thus responding to the so-called Benacerraf's dilemma (1973) – according to which the semantic and the epistemological *desiderata* are inconsistent in realism and anti-realism about objects.[11] The task of reconciling the

[11] On the one hand, ontological realism embraces a convincing semantics, whereby mathematical statements are interpreted at face value. Still, the abstract nature of these objects – which makes them not located in space and time and not causally effective – introduces serious epistemological

epistemological and the semantical requirements is performed by providing a more precise definition of mathematical structures and the positions within them.

The core of Shapiro's (1997) view consists of the label *ante rem* that characterizes this form of structuralism: structures exist *independently of* and *prior to* the systems of numbers which exemplify them – similarly to scientific OSR, *ante rem* structuralism is defined by the central role played by the notion of *structure*, that is both *fundamental* and *ontologically prior* to objects.

The reference to background structures allows mathematical statements to be interpreted at *face value*, since they do not generalize over all systems of objects but exactly refer to the positions within a specific structure, i.e. the natural numbers structure. This distinguishes *ante rem* structuralism (that is a form of realism, or *non-eliminativism* about structures) from eliminative structuralism (Benacerraf 1965; Hellman 1989), in which mathematical statements are generalizations over *all* systems of objects and the reference to both particular objects and the abstract structure is eliminated.

In an *ante rem* framework, mathematical objects are treated as *sui generis* entities, which are reduced to empty places or mere positions. This idea is captured by Shapiro's (1997) 'places-as-objects' perspective, in which (empty) places are legitimate objects in themselves, denoted by singular terms with their own properties and relations. Places as objects are completely determined for their identity by the structure they belong to. According to Shapiro (p. 6) to say that the number 2 is the second position in a particular progression (i.e the *ante rem* progression of the numbers) suffices to characterize it as a completely determinate entity:

> Roughly speaking, the essence of a natural number is the relations it has with other natural numbers. There is no more to being the natural number 2 than being the successor of the successor of 0, the predecessor of 3, the first prime, and so on.

For this reason, Shapiro's account appears significantly comparable with scientific OSR; on both accounts, all that matters about objects (i.e. quantum particles and numbers) are their *structural properties*.[12] As illustrated in Sect. 7.2, this assumption has led to the 'relation without *relata* objection' in OSR; when it comes to *ante rem* structuralism, the attempt of defining numbers in purely structural terms results in two similar worries:

(a) *'the problem of objects'*: to which extent are positions in a structure legitimate objects in themselves? In fact, places as objects appear to be too structurally defined to avoid resulting in a position where there are no objects (or even acceptable *entities*) at all.

problems. On the other hand, anti-realism about objects ensures a more straightforward epistemology but cannot account for a corresponding semantics, which would require objects to refer to.

[12] Shapiro (2006, 2008) has more recently developed a more moderate position about objects, according to which they possess non-structural properties as well (the property of being abstract, non-spatio-temporal, of not entering in any causal relation, etc., 2006, p. 116). Still, he has not really developed a view that accounts for mathematical objects in these terms.

(b) *'the problem of identity'*: on a structuralist conception of objects, structurally indiscernible objects are to be numerically identified with each other, in contrast with the mathematical practice.[13]

Concerning the 'problem of objects', Parsons (2008, p. 107) has sustained that «it is possible to have genuine reference to objects even if the 'objects' are impoverished in the way in which elements of mathematical structures appear to be».[14] Hellman (2001) and MacBride (2006) have developed more specific objections, intended to show an alleged circularity in the individuation of objects in *ante rem* structuralism: even though the identity of objects depends upon the relevant structures, structures presuppose *relata* having already been individuated or numerically distinguished.

Let us now consider the 'problem of identity' (Burgess 1999; Keränen 2001), related to the debate on whether the Principle of Identity of Indiscernibles (PII) can be maintained within the structuralist ontologies. This issue specifically emerges in non-rigid structures that allow for non-trivial automorphisms.[15] Such structures are composed by *distinct mathematical objects* that – if interpreted as mere positions, in accordance with Shapiro (1997) – turn out to be *structurally indiscernible* (for instance, $+1$ and -1 in the relative numbers structure and $+i$ and $-i$ in the complex numbers structure).

Several solutions have been proposed in the literature; some of them treat identity as a primitive notion (Button 2006; Ladyman and Leitgeb 2008; Shapiro 2008; Ketland 2011; Menzel 2018), arguing that the identity and diversity of places in a structure is accounted for by the structure itself. According to this solution, *ante rem* structuralism is not compelled to accept some versions of PII and is consistent with the mathematical practice, which sometimes concedes that indiscernible objects may be distinct; so, it is a mathematical fact that $+1$ and -1 are distinct, despite being indistinguishable within the structure. Other strategies introduce weaker forms of PII to deal with in a structuralist context; Ladyman (2005) – along the lines of Saunders'(2003) proposal in scientific structuralism (Sect. 7.3) – claims that numbers in structures with non-trivial automorphism are *weakly discernible* in virtue of the irreflexive relations holding between them (for example, for $+1$ and -1 in the relative numbers structure, 'to be the additive inverse of'). On that view, one can state the non-identity of mathematical objects without violating PII (or, at least, violating just the stronger versions of PII, which demand for an absolute or relative discernibility of objects)

Either way, the existing proposals raise further issues, thus leaving room to other solutions. In fact, the first strategy is not completely convincing, as the notion of primitive identity is controversial in the structuralist literature. The second one recalls the controversial notion of *weak discernibility* which, according to MacBride

[13] See Leitgeb (2020, part B sec.1–3) for this useful distinction.

[14] This worry has been introduced in Russell (1903), Benacerraf (1965) and Kitcher (1983).

[15] Internal symmetries that are not identity mappings.

(2006) does not actually face the objection (irreflexive relations still presuppose the numerical diversity of objects).

In what follows, I will advocate an alternative two-step strategy to resist the main objections to *ante rem* structuralism: first, I will make the relation between objects and structure more precise, by referring to the formulations of *ante rem* structuralism in terms of dependence (Linnebo 2008) and grounding (Wigglesworth 2018). Second, I will show that grounding is a better candidate to introduce Weak Mathematical Structuralism (WMS) as a more moderate interpretation of Shapiro's account, in which both 'the problem of objects' and the 'problem of identity' are avoided. WMS is directly based on Weak Structural Realism (WSR) in the philosophy of science and its solution to the 'relation without *relata*' objection in OSR.

7.6 Dependence and Grounding in *ante rem* Structuralism

As in scientific structuralism, the relation between objects and structures in mathematical structuralism stands in need of further clarification. Ontological dependence and metaphysical grounding accomplish this task, by expressing the fundamentality and the priority claims at hand in *ante rem* structuralism. These notions play a twofold role: first, they spell out the distinction between mathematical platonism and *ante rem* structuralism – after all, both positions refer to abstract objects embedded in larger structures, but just the second deems objects less fundamental than structures. Second, ontological dependence and grounding can be seen as non-eliminative relations that – for their very metaphysical features – assume the existence of both structures and objects. The reasons for interpreting ontological dependence as a non-eliminative notion have been clarified by Wolff (2012) in the context of WSR (cf. Sect. 7.4). Like ontological dependence, grounding is typically taken to be a non-eliminative notion, in which both the *relata* of the relation (that, in structuralist claims, are assumed to be objects and structures respectively) should exist: to say that A is grounded in B means than A obtains because of B, and not that is to be eliminated. This motivates the idea that grounded facts do not reduce to the facts grounding them:[16]

> [...]grounded facts and ungrounded facts are equally real, and grounded facts are an "addition of being" over and above the facts in which they are grounded' (Audi 2012, pp. 101–102).

[16] This means to reject the grounding-reduction link formulated by Rosen (2010), according to which if p reduces to q, then q grounds p: assuming grounding as a reductive notion would commit to an *identity* relation between the two facts – and not just to the claim that the grounded facts are less fundamental than or ontologically secondary to the *groundees* grounding them, which is the view here proposed.

Let us then evaluate how each notion works when applied to *ante rem* structuralism.

Linnebo (2008) refers to Lowe's (1994, 2016) identity dependence (ID) and introduces two different dependence claims:[17]

1. ODO (Objects Depend on Objects): each object depends for its identity upon all the other objects in the same structure.
2. ODS (Objects Depend on Structures): each object depends for its identity on the structure it belongs to.

These claims lead to a compromise view according to which certain mathematical objects depend on structures (e.g. abstract offices of the algebraic structures)[18] but not others (e.g. sets). Moreover, Linnebo (p. 78) distinguishes between a *strong* and a *weak* sense of dependence:

strong dependence: «*x* strongly depends on *y* just in case any individuation of *x* must proceed via y».

In accordance with Lowe (2003), 'individuation' means the *explanation* of the identity of an object. Applied to sets, strong dependence entails that in order to individuate a set, its elements must be specified – it is impossible to individuate a set (e.g. the singleton of Socrates) without proceeding *via* the individuation of its elements (e.g. Socrates himself). The reference to Lowe's (1994, 2016) (ID) makes strong dependence more precise:

> Since it is essential to the singleton of Socrates that it is the value of the singleton function applied to Socrates as argument, this singleton depends on Socrates. But since it is not essential to Socrates that he is the value of the sole-element-of function applied to the singleton as argument, there is no dependency in the reverse direction. (Linnebo 2008, *ibid*.).

However, there is another, 'weak' sense of dependence which, according to Linnebo (2008, *ibid*.), «has received little or no attention» in the literature:

weak dependence: «*x* weakly depends on *y* just in case any individuation of *x* must make use of entities which also individuate y».

For example, a set *weakly* depends upon its subsets; this is because it strongly depends on its elements, which also suffice to individuate the set's subsets.

Significantly, sets (which provide a counterexample to the dependence claim) do not *even weakly* depend upon their hierarchical structure of sets, whereas abstract

[17] Both claims are implicitly presupposed by Shapiro (2000, p. 253): «the number 2 is no more and no less than the second position in the natural number structure; and 6 is the sixth position. Neither of them has any independence from the structure in which they are positions, and as positions in this structure, neither number is independent of the other».

[18] Linnebo refers to 'abstract offices' in algebraic structures as places in structures obtained by a process of Dedekind abstraction, mapping a system to its abstract structure. While in a system offices can be filled by different sorts of occupants, the corresponding abstract structure «is left with nothing but the offices themselves» (Linnebo 2008, p. 75).

offices (to which dependence applies) depend *only weakly* upon algebraic structures. In particular, the individuation of an abstract office proceeds *via* (i.e. strongly depends upon) an ordered pair (R, *x*), where R is a system that realizes a structure and *x* an element in this system. It follows that an abstract office weakly depends on the other offices and on the abstract structure itself, «for in order to individuate such an office we need a realization of the structure. But this is also all we need to individuate the relevant abstract structure itself». (Linnebo 2008, p. 79).

As such, Linnebo's compromise view favors a non-eliminative approach to objects. However, Wigglesworth (2018, p. 223) has objected that the proposed account of dependence turns out to be not available to *ante rem* structuralism – to which Linnebo implicitly restricts his metaphysical investigation; in fact, according to Linnebo's definition of strong dependence, the individuation of abstract structures must proceed *via* a realization R. Arguably, R does not refer to a *particular* system (for any other system exemplifying the relevant structure would suffice to individuate it); what it is required is that *some* systems realize such structure. Even if this is the case, it follows that abstract structures strongly depend on the existence of some systems exemplifying them; but this is the thesis that is actually rejected by *ante rem* structuralism and endorsed by *in re* structuralism.

By contrast, Wigglesworth's (2018) interpretation of *ante rem* structuralism in terms of metaphysical grounding is supposed to supply a broader account, which relies on Linnebo's (2008) characterization of dependence and yet is consistent with an *ante rem* individuation of structures. The introduction of grounding relies on Linnebo's idea that the individuation of an object involves the explanation of its identity. Insofar the explanation at hand is *metaphysical explanation*, the dependence claims also qualify as grounding claims – given the close relation between dependence, grounding and metaphysical explanation.

Despite the analogies between grounding and dependence, grounding is generally understood as a distinct metaphysical relation, which fulfills stricter conditions (irreflexivity, asymmetry and transitivity). At its core, grounding captures the idea that some things obtain *because* or in *virtue of* some other things. If dependence holds between *entities*, the *relata* of the grounding relation are typically *fact*s or *propositions*. In particular, a fact is said to be grounded in another fact either for its *identity* or for its *existence*. In mathematical structuralism, grounding claims plausibly involve the *identity* of facts: «the fact that one entity has the identity it has grounds the fact that another entity has the identity it has». (Wigglesworth 2018, p. 225).[19] Another important distinction is that between full and partial ground.[20] Full ground entails that X on its own fully grounds Y; partial ground is generally defined in terms of full ground: X partially grounds Y just in case there is something else together with X such that they jointly ground Y.

[19] Shapiro (2008, p. 302) himself has rejected a form of existential dependence, given that mathematical objects necessarily exist.

[20] This distinction has been introduced by Fine (2012).

In such grounding framework, *ante rem* structures are identified with *unlabelled graphs* (G) composed by nodes (*n*) and edges (E) between the nodes – corresponding to objects and relations respectively – where E_n is the collection of the structural relations that a node instantiates and **G** is the isomorphism class of G.

With these clarifications at hand, let us investigate grounding claims in mathematical structuralism more deeply. In analogy with Linnebo's analysis (2008), two grounding claims are set out:

1. (ODO): for any mathematical objects, n_1 and n_2, in the structure G, the fact that the identity of n_1 is E_{n1} *partially* grounds the fact that the identity of n_2 is E_{n2}.
2. (ODS): For any mathematical object, *n*, in the structure G, the fact that G ∈ **G** *fully* grounds the fact that the identity of *n* is E_n.

The comparison between mathematical structuralism and graph theory allows to delineate straightforwardly identity criteria for structures: Wigglesworth argues that structures are not grounded for their identity in the nodes – which can be permuted leaving the graph unchanged – but in the operation of adding or removing an edge between the nodes, which would result in a different graph. This allows for an interpretation of grounding claims in terms of possible structures/graphs, which do not refer to any realization of the structure: «and so, unlike Linnebo's account, it is an account of grounding that is available to both the *ante rem* and *in re* non-eliminativist structuralist» (Wigglesworth 2018, p. 232). In a nutshell, the identity of a graph G is determined by its isomorphism class **G**. This is a standard definition of structures provided by Shapiro (1997, p. 93) in the context of *ante rem* structuralism:

> We stipulate that two structures are identical if they are isomorphic. There is little need to keep multiple isomorphic copies of the same structure in our structure ontology, even if we have lots of systems that exemplify each one.

Hence, Wigglesworth's (2018) account of grounding has the advantage of preserving an *ante rem* individuation of structures. A more detailed analysis of the properties of grounding provides further reasons to adopt grounding – rather than dependence – in order to account for (non-eliminative) structuralist claims. Even if both grounding and dependence are forms of metaphysical explanation, grounding is taken to have a stricter connection with metaphysical explanation, allowing in some cases for an identification of the two notions:[21] in fact, grounding, by being irreflexive, admits a full overlap with explanation, which standardly entails irreflexivity.[22] This fits well with the structuralist idea that structures ground objects in the sense of metaphysically explaining their identity, and reinforces the priority of structures by securing their explanatory import in mathematical structuralism. In terms of metaphysical explanation, the identity of an object is *partially* explained by its relations with any other objects in the same structure (ODO) and *fully* explained by the structure – there is *nothing outside the structure* explaining its identity (ODS).

[21] Among others, Dasgupta (2014) Raven (2015) and Thompson (2018) identify the two notions.

[22] This is not the case for dependence, which can be reflexive (i.e. an entity ontologically depends on itself).

On this view, grounding seems to capture the relation between objects and structures more deeply, and it is also a better tool to formulate WMS, which is the objective of the next section.

7.7 Weak Mathematical Structuralism (WMS) and Numbers as Thin Mathematical Objects

Let us briefly come back to scientific Weak Structural Realism, which asymmetrically defines the relation between objects and structures and appeals to Lowe's identity dependence (ID).

WSR: The identity of the putative objects/nodes is (asymmetrically) dependent (ID) on that of the relations of the structure.

In analogy with WSR, let us formulate WMS in terms of grounding:

WMS: The fact that an object has the identity it has is *fully* (and asymmetrically) grounded in the fact that the structure it belongs to has the identity it has.

The relevant sense of grounding at play is adequately captured by the (ODS) grounding claim, that addresses the asymmetrical relation between objects and structures:

ODS: for any mathematical object, *n*, in the structure G, the fact that G ∈ **G** *fully* grounds the fact that the identity of *n* is E_n.

By contrast, I will leave aside (ODO), which accounts for the symmetrical interdependence among objects and has been largely considered as a circular and not well-founded claim.

Articulating WMS in terms of (ODS), it seems that a category of objects can be individuated, which I will call *thin mathematical objects*: such reconceptualization of objects provides a variation of Shapiro's 'places as objects' and requires a significant reconsideration of their structural and non-structural properties. The comparison between mathematical structures and unlabelled graphs (cf. Ladyman and Leitgeb 2008; Wigglesworth 2018; Leitgeb 2020, part B) allows grasping thin mathematical objects more in detail. Within structures/graphs, objects can be understood as *unlabelled* and *edgeless* nodes, as illustrated in the following Fig. 7.2:

In my account, these nodes seem comparable to quantum particles in entanglement states (cf. fig. 1, p. 9). They are interchangeable because they can be permuted while leaving the graph unchanged; hence, their identity as individuals is solely determined by the relevant graph G'. However, the nodes in question cannot collapse

Fig. 7.2 Two unlabelled nodes in an edgeless graph

into one another, since they would result in a different (smaller) graph. Exactly as quantum particles, they appear as numerically distinguished *relata* whose existence results in both structural and non-structural properties.

In line with Linnebo (2008), structural properties in mathematical structuralism can be described as the properties that can be inferred through a process of abstraction (e.g. Dedekind's abstraction) or, similarly, as the properties that are shared by every system that instantiates the structures.[23] Still, this definition is subject to different counter-examples, concerned with non-structural properties of objects. Linnebo (2008, p. 64) takes into account the following cases:

> The number 8 has the property of being my favourite number. It also has the property of being the number of books on one of my shelves. And it has non-structural properties such as being abstract and being a natural number. In fact, the property of being abstract seems to be a very important property of natural numbers.

Here, different non-structural properties are mentioned: intentional properties (e.g. "being my favourite number"), applied properties (e.g. "being the number of books on one of my shelves"), metaphysical properties (e.g. "being abstract") and kind properties (e.g. "being a natural number").

According to Linnebo, to have non-structural properties is not equivalent to have *intrinsic properties*, defined as the properties that express the internal composition of objects, or the properties which an object would have «even if the rest of the universe were removed or disregarded» (Linnebo 2018, pp. 65–66).[24] As I have suggested, in the context of scientific structuralism non-structural properties of quantum particles plausibly qualify as kind properties (i.e. state-independent properties such as mass and spin). This interpretation may work also for mathematical structuralism, where thin mathematical objects are endowed with kind properties such as "being a natural number".

Very few attempts to clarify kind properties in the mathematical domain have been proposed. Intuitively, kind properties of numbers are strictly connected to their counting and measurement use in applicative situations: for example, natural numbers respond to the question "how many *Fs* are there?", whereas the rationals are defined for their role in measurement, i.e. as ratios between pairs of magnitudes, the reals as limits of Cauchy sequences of rationals, etc.[25]

Kind properties so understood are clearly non-structural (counting collections and measuring quantities are structure-independent operations) and yet non-intrinsic (they express the applicative function of numbers, and not their internal composition.). On the other hand, if we assume that places from different structures are distinct – as it is standard in *ante rem* structuralism – kind properties such as 'being a natural number' result in *essential* properties of numbers. This raises a

[23] See Schiemer and Korbmacher (2017) for a distinction between Linnebo's invariance account and Shapiro's (2008) definability account of structural properties in *ante rem* structuralism.

[24] Cf. Esfeld and Lam's (2011, p. 144) definition in scientific WSR (Sect. 7.4).

[25] This understanding of kind properties presupposes to interpret numbers as cardinals, rather than as ordinals.

possible objection to WMS (and, more broadly, to *ante rem* structuralism, whose typical slogan is that all the essential properties of numbers are structural). In this context, I will follow Shapiro (2006, p. 121) who acknowledges that while some extra-structural properties can be essential (such as 'being abstract', 'being non-spatio-temporal'), each property of a mathematical object 'comes in virtue of' its being the place it is in the structure to which it belongs. This plausibly generalizes to non-structural kind properties: the fact that the naturals, relatives, rationals, etc. are distinct – and then exhibit different kind properties – stems from their being 'tied' to different structures. As such, kind properties of numbers appear to be *secondary* to their structural properties and then can be accommodated in a structuralist account of objects.

On this basis, thin mathematical objects are defined by the conjunction of the following two conditions:

1.b) Thin objects [Identity]: thin objects are entities whose identity is entirely grounded in the structure.
2.b) Thin objects [Existence]: thin objects are entities whose *existence* (resulting in both structural and non-structural kind properties) is necessary to posit relations themselves.

I will now turn to the metaphysical issues (i–ii) presented so far in the philosophy of science, that concern thin mathematical objects as well.

(i) are thin mathematical objects substantial enough to avoid resulting in a 'no-objects-at-all' position?
(ii) are thin mathematical objects weak enough to preserve a structuralist framework?

I will face the first issue by investigating how thin mathematical objects respond to 'the problem of objects' and 'the problem of identity' in *ante rem* structuralism.

Let us start by the 'problem of objects'. Shapiro's places as objects are generally described as possessing structural properties only. By contrast, definitions 1.b and 2.b allow to elaborate numbers as *entities* which are endowed with both structural and non-structural kind properties. Such properties cannot individuate objects of the same kind, but are able to introduce them as numerically distinguished *relata*, conceivable independently of the structures and existing metaphysically prior to them. On that view, mathematical objects are not 'impoverished' in the way places as objects seem to be: there is actually *something* standing in the relations that are supposed to confer individuality on the *relata*, as required by the reference to (non-eliminative) grounding. So, a possible solution to the 'problem of objects' is put forward.

Significantly, this seems to apply to some cases of non-trivial automorphisms, which are relevant for the identity problem: consider, for instance, the relative numbers structure, in which the numbers $+1$ and -1 are discernible because $+1$ corresponds to the natural numbers kind, that is a subset of the relative numbers kind. Admittedly, $+1$ has the specific kind properties of the natural numbers, i.e. those properties which are used to count collections of objects. The negative relative

number −1 displays a different set of kind properties, extended from those of natural numbers in order to count collections in which negative quantities come into play (the counting of collections with just one/two . . . individual(s)). This solution is not unproblematic, and it is questionable whether it can be applied to other cases of non-trivial automorphisms. Nevertheless, it has the advantage of not involving either a primitive notion of identity or the reference to a weak form of PII, thus opening the path to a third way to overcome the identity problem.

Therefore, thin mathematical objects appear to be *substantial enough* to provide a possible response to the 'problem of objects' and 'the problem of identity', thus addressing the first issue introduced so far (*i.* are thin objects substantial enough?).

Let us now consider the second issue (*ii.* are thin objects weak enough?). First, thin mathematical objects lack intrinsic properties and any internal nature – their identity is solely given by the structure; in fact, admitting objects with intrinsic properties would be inconsistent with an *ante rem* structuralist framework,[26] and would rather commit to a platonist view about objects.[27] Nevertheless, one may object that 'being thin' could qualify as an intrinsic property itself. Recall Linnebo's definition of intrinsic properties as the properties which an object would have «even if the rest of the universe were removed or disregarded». This does not seem the case for the property of 'being thin', whose characterization relies on the conjunction of the two conditions Thin Objects [Identity] (1.b) and Thin Objects [Existence] (2.b): while the existence of a mathematical object – being necessary – could be in principle assumed independently of any other entities, this does not apply to its identity, which requires structural relations to be determined. So, given that both conditions should be fulfilled in order for an object to be 'thin', 'being thin' cannot be considered an intrinsic property in Linnebo's (2008) sense.

Second, thin objects do not commit to a form of *in re* structuralism, according to which abstract structures depend on the systems instantiating them. Conversely, they can be framed in an *ante rem* individuation of structures. As specified by Wigglesworth (2018), the reference to grounding enables us to ground the identity of structures/graphs in their isomorphism classes, where no systems are at play.

Hence, thin mathematical objects, though substantial enough to be legitimate *relata* of structural relations, are also *weak enough* to retain the priority of structures – in accordance with the asymmetry of WMS.

To sum up, thin mathematical objects appear as something more than Shapiro's mere positions but something less than the thicker objects occupying these positions in systems.

[26] In principle, *ante rem* structuralism is not inconsistent with platonism about objects (one can be committed to a background ontology of self-standing structures and yet admit objects, i.e. the natural numbers, which possess intrinsic properties and exemplify a specific structure); however, the same position appears quite odd if applied to Shapiro's places as objects which – by definition – have no more than their structural relations.

[27] Moreover, it is worth noting that intrinsic properties are consistent with other structuralist views, i.e. set-theoretic structuralism (in which sets have intrinsic properties of membership, making them 'self-standing') and *in re* structuralism (where any possible object can belong to the structures).

7.8 Concluding Remarks

Scientific Ontic Structural Realism (OSR) and mathematical *ante rem s*tructuralism are intimately related positions, which assume structures to be *fundamental* and *ontologically prior* and objects to be entirely reduced to their structural features. This entails that all that matters about objects are their structural properties, an assumption that has been deeply challenged in both the theoretical frameworks.

In the context of scientific structuralism, Weak Structural Realism (WSR) offers the most tenable form of OSR and successfully escapes the 'relation without *relata*' objection affecting eliminative views. WSR claims that objects (asymmetrically) depend on the structure for their identity, where the relevant sense of dependence – distinctively non-eliminative in character – is captured by Lowe's identity dependence. Though less fundamental than structures, quantum particles are reconceptualized as *thin physical objects*, endowed with both structural and non-structural *kind properties* (state-independent properties of quantum particles). Kind properties proved particularly useful to define thin physical objects (definitions 1.a and 2.a) and grasp them more specifically: on the one hand, they are something more than the 'no objects at all' perspective embraced by eliminative OSR, thus responding to the 'relation without *relata*' objection. On the other hand, they are something less than the thicker objects underlying the standard object-oriented metaphysics, so that structures remain ontologically prior.

A similar path has been explored in the mathematical framework, in order to deal with the 'problem of objects' and 'the problem of identity' in Shapiro's (1997) *ante rem* structuralism.

To this aim, I introduced Weak Mathematical Structuralism (WMS) as a novel position, elaborated in close analogy with WSR. WMS relies on (non-eliminative) metaphysical grounding and states that the fact that an object has the identity it has is fully (and asymmetrically) grounded in the fact that the structure has the identity it has.

The core of WMS consists of a significant reconsideration of Shapiro's interpretation of objects in terms of *thin mathematical objects*, understood as *unlabelled* and *edgeless nodes* in a graph. Similar to thin physical objects, such objects possess both structural and non-structural *kind properties*. In mathematical structuralism, kind properties of numbers turned out to be involved in counting and measurement facts, highlighting the different applicative uses of the naturals, relatives, rationals, etc. Two definitions of thin mathematical objects have been set out (definitions 1.b and 2.b) and a possible strategy to avoid the problem of objects and the problem of identity has been advocated. At the same time, thin mathematical objects are consistent with an *ante rem* individuation of structures. As such, thin mathematical objects are something more than mere positions but something less than the 'thicker objects' occupying these positions in systems.

In conclusion, WMS is advanced as a more moderate interpretation of *ante rem* structuralism, which attempts overcoming its main difficulties (i.e. the problem of objects and the problem of identity) without abandoning its core intuition (i.e. the priority of abstract structures).

References

Audi, P. 2012. Grounding: Toward a theory of the in-virtue-of relation. *Journal of Philosophy* 109: 685–711.

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

Burgess, J. 1999. Review of Stewart Shapiro (1997). *Notre Dame Journal of Formal Logic* 40: 283–291.

Busch, J. 2003. What structures could not be. *International Studies in the Philosophy of Science* 17: 211–225.

Button, T. 2006. Realistic structuralism's identity crisis: A hybrid solution. *Analysis* 66: 216–222.

Cao, T. 2003. Can we dissolve physical entities into mathematical structures? In Symonds (Ed.), Special issue: Structural realism and quantum field theory. *Synthese* 136 (1): 57–71.

Castellani, E. 1998. Galilean particles: An example of constitution of objects. In *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, ed. E. Castellani, 181–194. Princeton: Princeton Uni- versity Press.

Chakravartty, A. 1998. Semirealism. *Studies in History and Philosophy of Modern Science* 29: 391–408.

Dasgupta, S. 2014. On the plurality of grounds. *Philosopher's Imprint* 14 (20): 1–28.

Dorato, M. 2000. Substantivalism, relationalism and structural spacetime realism. *Foundations of Physics* 30 (10): 1605–1628.

Esfeld, M. 2004. Quantum entanglement and a metaphysics of relations. *Studies in the History of Philosophy of Physics* 35B: 601–617.

Esfeld, M., and V. Lam. 2008. Moderate structural realism about space-time. *Synthese* 160: 27–46.

———. 2011. Ontic structural realism as a metaphysics of objects. In *Scientific Structuralism*, ed. A. Bokulich and P. Bokulich, 143–160. Springer.

Fine, K. 1995. Ontological dependence. *Proceedings of the Aristotelian Society* 95: 269–269.

———. 2012. Guide to ground. In *Metaphysical Grounding: Understanding the Structure of Reality*, ed. Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press.

French, S. 2010. The interdependence of structure, objects and dependence. *Synthese* 175: 89–109.

———. 2014. *The Structure of the World: Metaphysics and Representation*. Oxford: Oxford University Press.

French, S., and D. Krause. 2006. *Identity in Physics: A Formal, Historical and Philosophical Approach*. Oxford: Oxford University Press.

French, S., and J. Ladyman. 2003. Remodelling structural realism: Quantum physics and the metaphysics of structure. *Synthese* 136: 31–56.

Hellman, G. 1989. *Mathematics Without Numbers: Towards a Modal Structural Interpretation*. Oxford: Oxford University Press.

———. 2001. Three varieties of mathematical structuralism. *Philosophia Mathematica* 9 (3): 184–211.

Hellman, G., and S. Shapiro. 2019. *Mathematical Structuralism*. Cambridge: Cambridge University Press.

Keränen, J. 2001. The identity problem for realist structuralism. *Philosophia Mathematica (III)* 9: 308–330.

Ketland, J. 2011. Identity and indiscernibility. *The Review of Symbolic Logic* 4 (2): 171–185. https://doi.org/10.1017/S1755020310000328.

Kitcher, P. 1983. *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.

Ladyman, J. 1998. What is structural realism? *Studies in History and Philosophy of Science* 29: 409–424.

———. 2005. Mathematical structuralism and the identity of indiscernibles. *Analysis* 65: 218–221.

———. 2007. On the identity and diversity of individuals. *The Proceedings of the Aristotelian Society* 81: 23–43.

———. 2020. Structural realism. In *Stanford Encyclopedia of Philosophy,* ed. E.N Zalta, https://plato.stanford.edu/archives/spr2020/entries/structural-realism/.
Ladyman, L., and Leitgeb. 2008. Criteria of identity and structuralist ontology. *Philosophia Mathematica* 16: 388–396.
Ladyman, J., and D. Ross. 2007. *Every Thing Must Go: Metaphysics Naturalised*. Oxford: Oxford University Press.
Leitgeb, H. 2020. On non-eliminative structuralism: unlabeled graphs as a case study: Part B. *Philosophia Mathematica*. https://doi.org/10.1093/philmat/nkaa009.
Linnebo, Ø. 2008. Structuralism and the notion of dependence. *Philosophical Quarterly* 58 (230): 59–79.
Lowe, E.J. 1994. Ontological dependency. *Philosophical Papers* 23 (1): 31–48.
———. 2003. Individuation. In *Oxford Handbook of Metaphysics*, ed. M. Loux and D. Zimmerman, 75–95. Oxford: Oxford University Press.
———. 2005/2010. Ontological dependence. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta, https://plato.stanford.edu/archives/spr2010/entries/dependence-ontological/.
———. 2012. Asymmetrical dependence in individuation. In *Metaphysical Grounding*, ed. F. Correia and P. Schnieder, 214–233. Cambridge: Cambridge University Press.
MacBride, F. 2006. What constitutes the numerical diversity of mathematical objects? *Analysis* 66 (1): 63–69.
McKenzie, K. forthcoming. Structuralism in the idiom of determination. *British Journal for the Philosophy of Science*: axx061.
Menzel, C. 2018. Haecceities and mathematical structuralism. *Philosophia Mathematica (III)* 2: 84–111.
Morganti, M. 2004. On the preferability of epistemic structural realism. *Synthese* 142: 81–107.
Parsons, C. 2008. *Mathematical Thought and its Objects*. Cambridge: Cambridge University Press.
Psillos, S. 2001. Is structural realism possible? *Philosophy of Science* 68 (Supplementary Volume): S13–S24.
Quine, W.V. 1960. *Word and Object*. Harvard: Harvard University Press.
Raven, M.J. 2015. Ground. *Philosophy Compass* 10 (5): 322–333.
Rosen, G. 2010. Metaphysical dependence: grounding and reduction. In *Modality: Metaphysics, Logic, and Epistemology*, ed. B. Hale and A. Hoffman, 109–136. Oxford: Oxford University Press.
Russell, B. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.
Saunders, S. 2003. Physics and Leibniz's principles. In *Symmetries in Physics: Philosophical Reflection*, ed. K. Branding and E. Castellani, 289–307. Cambridge: Cambridge University Press.
Schiemer, G., and J. Korbmacher. 2017. What are structural properties? *Philosophia Mathematica (III)* 26: 295–323.
Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.
———. 2000. *Thinking About Mathematics*. Oxford: Oxford University Press.
———. 2006. Structure and identity. In *Identity and Modality*, ed. F. MacBride, 34–69. Oxford: Oxford University Press.
———. 2008. Identity, indiscernibility, and ante rem structuralism: The tale of i and –i. *Philosophia Mathematica (III)* 16: 285–230.
Stachel, J. 2002. The Relations between things versus the things between relations. The deeper meaning of the hole argument. In *Reading Natural Philosophy: Essays in the History and Philosophy of Science and Mathematics*, ed. D. Malament, 231–266. Chicago/LaSalle: Open Court.
Stein, H. 1989. Yes, but . . . some skeptical remarks on realism and antirealism. *Dialectica* 43: 47–65.
Tahko, T., and E.J. Lowe. 2016.Ontological dependence. In *Stanford Encyclopedia of Philosophy,* ed. E.N. Zalta., https://plato.stanford.edu/archives/win2016/entries/dependence-ontological/.

Thompson, N. 2018. Metaphysical interdependence, epistemic coherentism, and holistic explanation. In *Reality and Its Structure, Essays of Fundamentality*, ed. Ricki Bliss and Graham Priest, 107–126. Oxford: Oxford University Press.

Wigglesworth, J. 2018. Grounding in mathematical structuralism. In *Reality and Its Structure: Essays in Fundamentality*, ed. R. Bliss and G. Priest, 217–236. Oxford: Oxford University Press.

Wigner, E. 1939. On unitary representations of the inhomogeneous Lorentz group. *The Annals of Mathematics, Second Series* 40: 149–204.

Wolff, J. 2012. Do objects depend on structures? *British Journal for the Philosophy of Science* 63 (3): 607–625.

Chapter 8
The Structuralist Mathematical Style: Bourbaki as a Case Study

Jean-Pierre Marquis

Abstract In this paper, we look at Bourbaki's work as a case study for the notion of mathematical style. We argue that indeed Bourbaki exemplifies a mathematical style, namely the structuralist style.

Keywords Philosophy of mathematics · Epistemology · Philosophy of mathematical practice · Bourbaki · Mathematical style

8.1 Introduction

In his article in the *Stanford Encyclopedia of Mathematics* on mathematical style, Paolo Mancosu presents the challenge of developing an "epistemology of (mathematical) style":

> Are the stylistic elements present in mathematical discourse devoid of cognitive value and so only part of the coloring of mathematical discourse or can they be seen as more intimately related to its cognitive content? (Mancosu, 2017)

There is no doubt that there are stylistic elements in the *presentation* of mathematics. After all, writing and talking about mathematics is not purely a matter of manipulating formal symbols organized in a unique manner. It is another issue to

The author gratefully acknowledge the financial support of the SSHRC of Canada while this work was done. This paper is part of a larger project on Bourbaki and structuralism which would not have seen the light of day without Michael Makkai's influence and generosity. I want to thank him for the numerous discussions we had on the subject. I also want to thank the organizers of the FilMat conference which was held in beautiful Mussomeli. Finally, I want to thank Leo Corry, Robert Thomas and Elaine Landry whose comments and criticisms allowed me to move from a bad draft to what I hope is a coherent paper.

J.-P. Marquis (✉)
Département de Philosophie, Université de Montréal, Montréal, QC, Canada
e-mail: Jean-Pierre.Marquis@umontreal.ca

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_8

determine whether there are stylistic features in *mathematics*. Asking the question brings us immediately to the *practice* of mathematics and all its aspects. One has to define a concept. One has to state a theorem. One has to prove a theorem. One has to construct a counter example. One has to find a method to compute a formula. Etc. More often than one might think, in most cases, there is no unique path to the solution to a given problem. Then, one has to write. One has to talk. And there are myriad presentations possible. There are many different ways to introduce and justify a definition, motivate and contextualize a theorem, write up a proof and even organize a computation. Of course the plurality of presentations is not unique to mathematics. Anyone who has to present and prepare some material is faced with similar challenges. Is there an element of style that would be intrinsic to mathematics? Or, at least, a style that would bring an epistemological dimension that cannot be dissociated from the style? Are there styles of definitions, styles that contain an inherent epistemological component? This is how I understand Mancosu's challenge. Thus, if it is taken up, its resolution has at least two parts. First, identify what constitutes the stylistic elements in mathematical knowledge, as opposed to methods, approaches, etc., or merely "color". Second, show that these stylistic elements have cognitive value and, again, are not merely "part of the coloring".

It seems a priori easy to identify what the "coloring" of mathematical discourse might be: it should be some kind of ornament that accompanies a discourse, but that does not essentially contribute to its cognitive content. The terminology itself brings us back to the arts, any art, be it music, painting, sculpture, dance, acting, literature, etc. This is the traditional association. If mathematical style is merely part of the coloring, it would be akin to literary style, even a special case of the latter, it would refer to a specific way of writing a mathematical presentation, dictated by esthetic choices that do not have an impact on the epistemological content of the mathematics itself. It may make a mathematical text clearer, more fun, more powerful, more enjoyable or what have you, but if we are in the realm of coloring, then it does not convey a specific epistemic content, it does not contribute to its justification. It could be completely removed and the mathematics would be in principle just as clear, just as right, just as justified. Underlying this conception of mathematical knowledge is the idea that the truths of mathematics are organized in an essentially unique logical network and that to know mathematics is to know this web of logical relations. Whatever is added to this network would be ornemental, for instance the use of pictures, of certain types of notations and symbols and, of course, the presence of texts that are not directly part of the logical deductions. However, as Mancosu points out himself, it is easy to find claims in the philosophical literature that there *are* mathematical styles, be they individual styles, national styles or epistemic styles. I refer the reader to the list he provides in Mancosu (2017).

This paper is an attempt to face Mancosu's challenge head on[1] by examining the case of Nicolas Bourbaki, the well-known collective of French mathematicians that initiated an ambitious and influential undertaking in 1934 and that led to the publications of 28 volumes covering a large spectrum of modern mathematics—from set theory to algebraic topology.[2] The project was initially meant to provide a modern treatise on analysis, but quickly became something much larger, since Bourbaki decided to start from scratch and organize the material from an abstract standpoint. Bourbaki's work was extremely influential and contributed to the development of contemporary mathematics in many different ways. It had, in the 1960s, its supporters and detractors. Although the collective still officially exists and still organizes an important seminar held in Paris, what we will focus on in this paper is the work done by the first two generations of Bourbaki, namely the founding fathers and those who joined the collective after WWII. Our main claim is that *this* Bourbaki is a generic case of an *epistemic* mathematical style. We also claim that this style is a direct consequence of a very specific conception of mathematics, its nature, organization and articulation, namely the structuralist style. Given these goals, the plan of our paper is straightforward. We will first propose a general definition of mathematical style. Then we will take a close look at Bourbaki's mathematics. We will then step back and try to explain what we mean by the structuralist style.

8.2 The Notion of Mathematical Style

As a first approximation, I submit that a mathematical style is a *systematic way of doing* mathematics which is then represented in its presentations. More precisely, it is a global and systematic pattern of choices that are made to define concepts, prove or disprove theorems, solve problems, compute formulas. Note that we are within a set of goal-oriented activities. For this approximation not to be a platitude, I have to put some flesh around the bones. By systematic, I mean that the way of doing is repeated, thus is identifiable and used more than once. A style, be it mathematical or otherwise, cannot be a fluke, a singular manifestation of a behavior. It has to be a way of behaving, of doing, of making that is a variation or a series of variations around an identifiable pattern, even though the latter might be hard to define. But

[1] As I will make clear in the second section, I am not the first one to do so. One could go back to Gilles Gaston Granger's work on the notion in Granger (1968), also discussed in Mancosu's article. I will not use nor refer to Granger's work here, for it would take us too far from our main objective. I will, however, follow the steps of David Rabouin in Rabouin (2017).

[2] It is hard to determine the exact number of volumes, particularly in the original publications, since some were published in parts and then in complete volumes. The best source of information about the origins of the collective and its early work is still found in Beaulieu (1990). See also Beaulieu (1994). Other important sources are Corry (1992, 1996, 2001, 2009). Corry's work is extremely valuable and stimulating.

even for this to be possible, I claim that the following conditions have to be satisfied. In order to have a mathematical style, there has to be:

1. A "standard", a way of doing mathematics against which the alternative style is contrasted; most of the time, this standard is implicit and is not recognized as such by the practitioners;
2. A combination of patterns of behavior that deviate significantly from the standard;
3. A systematic and voluntary use of these patterns; these patterns have to be used and sought in all possible cases. They are not adopted as a mere option that can be discarded at will and they are not practiced without the practitioner being willingly aware of them; it has to be implemented as a conscious value in and for itself.

Some comments and clarifications are required. The first condition, namely the existence of an implicit (or explicit) standard seems to me to be inescapable. A style, to be identified as a style, has to be distinctive in one way or another and for this to be possible, there has to be something that it is distinguished from.

A style, to be recognized as a style, has to deviate significantly from a (implicit or explicit) standard. It is, of course, difficult to qualify in general what 'significantly' means precisely. Within a given practice, there are variations. These variations by themselves do not yield nor do they constitute a style, but they might the precursor to a style. The expression 'combination of patterns of behavior' refers to ways of doing that are guided or systematic, that follow a pattern. Of course, it is not a method nor a combination of methods in the sense of an algorithm or algorithms. For a style is fluid, changing within a certain range or space of variations, but also rigid enough so that it can be recognized as such.

Last but not least, for a style to be a style, the patterns of behavior underlying it have to be consciously adopted and applied in all possible cases, even those that might seem outside the original scope of these behaviors. One subtle point has to be made about the voluntary aspect. It is not that the person who is adopting these patterns is aware that she is adopting a style—for she might not think of it in these terms—,but she has to be aware that the patterns of behavior she is adopting *are* deviant from the standards of the community. It is entirely possible that she is simply adopting the patterns of behavior that seem to her to be the best or most effective given her goal, what she knows and what she can do. In other words, sometimes a style comes naturally and is not seen as being the result of a conscious effort to behave the way that person behaves. But in the eyes of others, it is definitely a style.

Notice that a style is intrinsically historical. It appears at some point and can, and usually does, disappear at another point. But it has to last sufficiently long so that it can be identified as such. It can also become the new standard for a given community and thus lose its status as a style if the new generations are taught to do mathematics by adopting these behavioral patterns.

We propose a more precise definition of mathematical style. Let us fix a few conventions. First, by an agent α, we refer to the author of a piece of mathematics, be it an individual, a group of individuals, a collective, etc. Second, by a cultural

context γ, we refer to the accepted norms, implicit or explicit, in a given community, that dictate how a certain activity has to be performed or is usually performed. Needless to say, there are specialized cultural contexts, e.g. homological algebra, descriptive set theory, etc., as well as more global cultural contexts, number theory, algebra, analysis, even mathematics as a whole. Third, by patterns of definitions δ, we refer to ways of using a language, spoken or written—we use the term 'language' in a broad sense, including diagrammatic, visual, symbolic, etc., conventions —, even introducing a new language and using it in a certain manner. Fourth, by patterns of inference ι, we refer to ways of arguing based on the choice of linguistic and/or symbolic devices made, that is the choice of δ. Contrast and compare, for instance, the way mathematicians can now define the product of two sets X and Y. In the language of set theory, one defines the Cartesian product in the usual fashion, that is as a set containing the ordered pairs (x, y), with $x \in X$ and $y \in Y$. In the language of category theory, a product of two sets is defined as being an object P together with two morphisms p_X and p_Y satisfying the usual universal property. These choices then determine to a certain extent the patterns of inference one can use and will use, even though there is still room for variations in the patterns of inferences employed within both contexts. We can now give our definition.

We say that a corpus of mathematics μ, embodied in books, papers, talks, etc., produced by an agent α exhibits or has an *epistemic style* σ in the cultural context γ if and only if σ is a systematic way of solving problems that rests upon:

i. specific and systematic patterns δ of definitions that differ significantly from the standards of γ;
ii. specific and systematic patterns ι of inference that differ significantly from the standards of γ;
iii. combinations κ of components of δ and ι in the solution of problems, the organization of concepts, results and relations between the parts of μ that differ significantly from the standards of γ.

This gives a general definition of a mathematical style, but it does not provide the features of a *particular* mathematical style. To get the latter, one has to define δ, ι, their combinations κ and specify how they deviate from the standards of γ.

Let me immediately illustrate this definition by a concrete example, which I hope will be useful. I claim that members of the contemporary community of logicians, mathematicians and computer scientists who are developing and using homotopy type theory could end up practicing a new style of mathematics in the foregoing sense. I cannot, of course, describe homotopy type theory in such a short paper.[3] I will sketch the main elements that I believe can justify my claim.

[3] See Collective (2013) for a presentation of the theory and how mathematics is developed within it. Of course, homotopy type theory is also presented as a new foundational framework. As a consequence, it is taken to be global and systematic, which are two elements that are crucial to our approach. I underline again that the 'global' is always relative to a community. It could be all of homological algebra, or all of algebraic geometry, but not all of mathematics, for instance.

First, homotopy type theory can be used systematically to do mathematics, to solve mathematical problems. It differs as such significantly from the standards used, implicitly or explicitly, by the contemporary community of mathematicians. The language of homotopy type theory is not based on the standard universe of sets or a variant thereof. Classical mathematical entities are defined by using new means of definitions and theorems and computations are obtained by novel inferential and computational patterns. Classical constructions and concepts, e.g. sets, the homotopy groups, the Hopf fibration, Eilenberg-Mac Lane spaces, etc., are defined in novel ways and proofs of theorems have to go through new paths (no pun intended). See Licata and Finster (2014), Rijke and Spitters (2015), and Buchholtz et al. (2018) for some examples.

I want to emphasize that we need not have such a well-defined, formal framework to characterize a singular mathematical style. In fact, as it develops, the mathematical practice based on homotopy type theory might become a mixture of purely formal, computational mathematics, checked by computers, and informal expositions containing the main mathematical ideas involved in the computations. However, the latter does not constitute its style. Its style resides in the patterns of definitions, patterns of inferences and their combinations in the solutions of mathematical problems. It is not tied up to specific axioms, e.g. the univalence axiom, of homotopy type theory, but rather basic methodological features built into it. Thus, some of the technical, formal aspects of homotopy type theory might be modified, even abandoned, and the style could still be present. The style is not attached to the specific (univalent) foundational framework presented and explored, but rather to the language, the manners of defining, proving and calculating that can be kept apart from the specific formal framework.

To make sure that our example does not mislead the reader in thinking that our definition of mathematical style applies only to formalized mathematics and formalized theories, let me fall back on a recent analysis of the notion of mathematical style proposed by David Rabouin. Even though Rabouin does not give a general definition of the notion of mathematical style in his paper (Rabouin, 2017), his approach is close to ours in many respects and has, in fact, inspired ours.[4]

Based on Chevalley's paper on mathematical style, Rabouin identifies a mathematical style with a way of writing that inflects mathematical thought. After pointing out that Chevalley does not give a definition of mathematical style, Rabouin presents Chevalley's position thus: "…he [Chevalley] merely states that one can identify general tendencies in ways of writing mathematics…"(Rabouin, 2017, 142), and then quotes Chevalley saying that there are "revolutions that inflect writing, and thus thought."(Rabouin, 2017, 145). There are other elements that are implicitly included in Rabouin's analysis. Two features have to be underlined, for they are directly tied to our analysis. The first component has to do with patterns of inferences, which he mentioned in an example:

[4] There are also parallels with (Kvasz, 2008), but we will not expand on this particular point here.

> When Poincaré used the ϵ-style, it was not because he shared a certain conception with Weierstrass of what the objects (...) involved in this manipulation were and about the good (in this case 'rigorous') delineation of the theories, *but because this way of writing allowed some powerful inferences that were not possible in the previous style*. (Rabouin, 2017, 148)[my emphasis]

Finally, another relevant element comes up when he discusses the Cartesian style:

> Both Descartes's and Fermat's methods rely on a kind of inferential black box coupled with geometrical reasoning. This allows us to give a more precise characterization of the Cartesian style (at least for one important aspect): its core is not the use of algebra in and of itself (which existed long before Descartes and Fermat) but *the coupling of specific kinds of computational inferences with geometrical ones*. In this sense, one can say that the Cartesian style of geometry, even if it did not suddenly disappear, took a dramatic turn around 1750 with the first formulations, which, as later emphasized by Joseph-Louis Lagrange, were free from any diagrammatic inferences—Leonhard Euler (1748) can be considered a starting point here. (Rabouin, 2017, 154)[my emphasis]

It is not only the computational inferences but also the geometrical inferences that we want to underline here, which we include in the patterns of inferences contained as an intrinsic part of the language or the writing.

Rabouin gives also the examples of Leibniz's style of (transcendental) geometry, set theory as a language (as opposed to formalized set theory) and the Euclidean style of geometry[5] as examples of his notion of mathematical style.

As emphasized by Rabouin, a style can be adopted for a variety of reasons, even incompatible reasons, and these reasons are not necessarily philosophical. For some, it might be associated with a specific ontology. To others, it might be seen as a consequence of a chosen epistemology. It is even conceivable that some see in it an ideological or political component. Finally, it might simply be more effective than another way of solving certain problems. The main point here is that the style is not *defined* by only a common ontology or a common epistemology, etc.

I will now try to show that Bourbaki is an exemplar of the notion of a mathematical style.[6]

8.3 Bourbaki's Style

Bourbaki is particularly interesting when looked at from the point of view of the notion of mathematical style. The fact is, we could use the expression "Bourbaki's style" in three different senses.

[5] At least as interpreted by Ken Manders in his Manders (2008).

[6] I am using the term 'exemplar' in a sense similar to Kuhn's usage in his postscript of the second edition of Kuhn (1970).

1. Bourbaki had a unique method of work; it was a collaborative effort unlike any other before and, as far as I know, ever since.[7] This in itself deserves to be called "Bourbaki's style of work".
2. Bourbaki developed a unique, terse way of presenting mathematics which even became known as "Bourbaki's style". We can therefore talk about "Bourbaki's presentation style".
3. Finally, and most importantly for our project, Bourbaki's modes of development of mathematics itself, centered on a certain notion of structure and of how to do mathematics in a structuralist fashion. It is of course at this level that our characterization of the notion of mathematical style ought to apply to Bourbaki. Thus, there is "Bourbaki's structuralist style".

These three senses of styles are not independent. The third, namely the mathematical style as such, emerged in part from the first, Bourbaki's method of work. The second, the writing style, is a direct consequence of the third and the first components. We will look at these three senses in turn. But before we do so, we have to provide a minimum amount of information about Bourbaki, for it is an essential part of the context.

8.3.1 Bourbaki: A Very Short Description of the Group and the Project

Bourbaki was famous among mathematicians, and intellectuals in general, from the 1960s until the beginning of this century approximately. The new generation of philosophers, logicians and mathematicians have very little knowledge of who they are, what they did and why it was important, and thus it seems appropriate to give a short presentation of the group.[8]

André Weil (1906–1998), Henri Cartan (1904–2008), Claude Chevalley (1909–1984), Jean Delsarte (1903–1968), Jean Dieudonné (1906–1992), René de Possel (1905–1974), a group of young and ambitious mathematicians, all former students from the *École Normale Supérieure* in Paris, an élite school, met for the first time in December 1934 to discuss the idea of writing together a modern textbook in analysis. Except for Claude Chevalley, the youngest member of the group, they were all university professors who found that they did not have at their disposal a decent textbook to work with and Weil convinced them that the best solution was simply

[7] The closest I can find nowadays are The Stack Project in algebraic geometry, the nLab in higher dimensional category theory and Gowers's Polymath Project. But they all differ in one way or another from Bourbaki's work. See https://stacks.math.columbia.edu/about for the Stack Project, https://ncatlab.org/nlab/show/HomePage for the nLab and https://en.wikipedia.org/wiki/Polymath_Project for the Polymath project.

[8] There is nothing original in this section. The interested reader can consult (Beaulieu, 1990, 1994; Corry, 2004, 2009; Houzel, 2004; Mashaal, 2000) for more.

to write one. They certainly did not know then that they had just set in motion a unique collaborative enterprise that would not only last well after their withdrawal from the group, but that would also have a deep impact on the face and development of mathematics in the twentieth century.

They were well aware that mathematics was changing and that Hilbert and his school were promoting the axiomatic method in mathematics. Many of them had visited Göttingen, Berlin, Hamburg, Frankfurt, Munich, Rome, Stockholm, Zurich, Copenhagen, Princeton (to mention but the most important places) during their graduate studies or afterwards. They had all read Van der Waerden's *Moderne Algebra* and it had a great impact on them.

The first "extensive" meeting took place in the summer of 1935.[9] The composition of the group changed somewhat in the meantime and would change again in the following fall. We will not follow the exact composition of the group through time. Suffice it to say that Weil, Cartan, Chevalley, Delsarte and Dieudonné formed the core of the group for the first 20 years or so. Charles Ehresmann (1905–1979) joined the group in the fall of 1935 and left in 1947. After WWII, Laurent Schwartz (1915–2002), Pierre Samuel (1921–2009), Roger Godement (1921–2016), Jean-Louis Koszul (1921–2018), Armand Borel (1923–2003), Jean-Pierre Serre (1926-), Alexandre Grothendieck (1928–2014) and Pierre Cartier (1932-) joined the collective at some point. Samuel Eilenberg (1913–1998), one of the fathers of category theory, became a member in 1950. All the members were creative mathematicians who all had respected individual careers. All of them nonetheless said that being members of Bourbaki and working together had a deep influence on their individual work.[10]

The original plan was simple enough: write a modern textbook on analysis. It became clear that they needed to start with what they called an "abstract packet", which included set theory, general topology and algebra as it was then known. Notice that these three disciplines were being created at the time. Indeed, Bourbaki contributed to their evolution and stabilisation.[11] What was supposed to be merely an introductory chapter rapidly became a large undertaking. Bourbaki first published a fascicle of results on set theory in 1939. It was not, as such, a textbook, for it contained no proofs. They decided to publish it nonetheless, since many of the results on sets were to be used in subsequent volumes. The complete volume on set theory finally came out in two parts, one published in 1954 and the other in 1957. The last one contains the chapter on structures. Notice how long it took them to finally get it published: almost 20 years. The volume on sets and structures has a

[9] In fact, the group had met every two weeks during the winter and the spring of 1935. The summer meeting was an intensive session where they hoped to do more work together.

[10] In a late interview, Henri Cartan declared: "In Bourbaki I learned very much. Almost all I know in mathematics I learned from and with the Bourbaki group."(Jackson, 1999, 785).

[11] Suffice it to say that the axioms of topology in terms of open sets came directly from Bourbaki. So does the notation for the empty set, $\emptyset$, among other things.

tortuous history and it went through numerous versions.[12] It might be worth pointing out that the *general* notion of structure was not in Bourbaki's mind in 1935. It showed up for the first time during the meeting held in the summer of 1936, but as an undefined concept. It then went through various presentations and the final, published version, did not satisfy the group, for reasons that we will clarify later.

Despite the fact that they had to modify the original project in numerous ways and even, at some point, to scale it down, Bourbaki started publishing books as early as 1940. The first volume contained the first chapters of *General Topology*, quickly followed by the first chapters of *Algebra* in 1942. Subsequent chapters on topology and algebra follow in 1947 (topological groups, linear algebra), 1948 (multilinear algebra, real numbers), 1949 (functions of a real variable, functional spaces) and in the 1950s, they basically published a volume a year, up to the theory of integration. It was an intensive undertaking, ambitious and systematic. No single author could have done that. Even for a distinguished group, and especially given their method of work, it is remarkable that they succeeded in doing anything.

8.3.2 Bourbaki's Method of Work

Team work is neither easy nor simple.[13] A large amount of trust and respect has to exist between the members for anything to be done. There also has to be an agreement as to what the final goal is, otherwise the group spends countless hours wasting time discussing that goal. In Bourbaki's case, the target was clear at the beginning, but it changed as the work developed. Somehow, the original members agreed on a method of work and it led to the publications mentioned. The method was brutal. Here is how Dieudonné presented it later.[14]

> The work method used in Bourbaki is a terribly long and painful one, but is almost imposed by the project itself. In our meetings, held two or three times a year, once we have more or less agreed on the necessity of doing of book or chapter on such and such a subject (...), the job of drafting it is put into the hands of the collaborator who wants to do it. So he writes one version of the proposed chapter or chapters from a rather vague plan. Here, generally, he is free to insert or neglect what he will, completely at his own risk and peril, ... After one or two years, when the work is done, it is brought before the Bourbaki Congress, where it is read aloud, not missing a single page. Each proof is examined, point by point, and criticized pitilessly. One has to see a Bourbaki Congress to realize the virulence of this criticism and how it surpasses by far any outside attack. (...) Once the first version has been torn to pieces – reduced to nothing – we pick a second collaborator to start it all over again. This poor man knows what will happen because although he sets off following new instructions,

[12] It is now possible to consult these versions on line, since the early documents have been digitized and made available on the site http://sites.mathdoc.fr/archives-bourbaki/

[13] For more on Bourbaki's method of work, the reader can consult the references given in the previous footnotes.

[14] Other original members have provided similar descriptions and later members concurred. See, for instance Guedj (1985) or Cartan (1979).

> meanwhile the ideas of the Congress will change and next year *his* version will be torn to bits. A third man will start, and so it will go on. One would think it was an endless process, a continual recurrence, but in fact, we stop for purely human reasons. When we have seen the same chapter come back six, seven, eight, or ten times, everybody is so sick of it that there is a unanimous vote to send it to press. This does not mean that it is perfect, and very often we realize that we were wrong, in spite of all the preliminary precautions, to start out on such and such a course. So we come up with different ideas in successive editions. But certainly the greatest difficulty is in the delivery of the first edition (Dieudonné, 1970, pp. 141–142).

The result was perhaps not perfect, but very few books are written in that way and go through such a rigorous editing process. Although not faultless, the final result was certainly better is some ways than what it would have been had it been the product of a single individual. Definitions were weighed, proofs were criticized, the organisation of theorems was analyzed, the overall network of concepts and results was evaluated by first-rate mathematicians. The result was something unique. There is one important component of the method that Dieudonné did not underline. As Chevalley later put it: "This allowed our work to submit to a rule of unanimity: anyone had the right to impose a veto. As a general rule, unanimity over a text only appeared at the end of seven or eight successive drafts."(Guedj, 1985, 47). Majority was not enough. If only one member thought that a manuscript was not good enough, it had to be rewritten. Like I said, it was a brutal process.

This mode of collaboration certainly played a role in the redactions of the volumes published over the years. It contributed in an essential way to the construction of the presentation of the material and its organization. For, when one looks at the works of its individual members, it is clear that there are differences between what Bourbaki published and what they published, even when some of the members produced expository material. Chevalley, for instance, is more radical than Bourbaki in some ways. I cannot refrain from quoting a long passage from a review of Chevalley's textbook on algebra, (Chevalley, 1956), written by Mattuck:

> Chevalley has written a text-book, and his mathematical personality permeates every paragraph. [...] The book is tight, unified, direct, severe; relentlessly and uncompromisingly it pursues its ends: out of the simplest basic notions of algebra to build up with perfect precision the theory of multilinear algebras which have found applications in topology and differential geometry. [...] The unity is monolithic. Gone is the discursive rambling of previous texts. This one marches unswerving and to its own music. [...]
>
> The general approach to the subject matter is that of Bourbaki's first three algebra chapters, but there are significant differences in content and treatment (Chevalley is often more general). *As for the style, Bourbaki emerges from the comparison a warm, compassionate, and somewhat elderly gentleman*. (Mattuck, 1957, 412)[my emphasis]

Mattuck directly refers to Bourbaki's style or presentation and compares it to Chevalley's. There is no doubt that to characterize Bourbaki's style of presentation as "warm, compassionate, and somewhat elderly" was deeply ironical. To most readers at the time, Bourbaki was anything but warm, compassionate and somewhat elderly! Mattuck's description of Chevalley's book as "tight, unified, direct, severe" is precisely what its contemporaries would have claimed of Bourbaki's books. Chevalley was pushing it even further.

Weil, on the other hand, wrote books that are definitely not in Bourbaki's style, at least not in the sense that I am using the term. For instance, in his review of Weil's *Foundations of Algebraic Geometry*, Oscar Zariski underlines the fact that "It is a remarkable feature of the book that—with one exception (Chap. III)—no use is made of the higher methods of modern algebra. The author has made up his mind not to assume or use modern algebra 'beyond the simplest facts about abstract fields and their extensions and the bare rudiments of the theory of ideals'."(Zariski, 1948, 671). Zariski himself claims afterwards that "we may just as well help ourselves to modern algebra to the fullest possible extent", a claim certainly consistent with Bourbaki's style. And then, he goes on, this time talking about Weil's writing itself: "To achieve his objectives Weil wages a campaign of the Satz-Beweis type. Most readers will find it difficult to follow the author through the seemingly endless series of propositions, theorems, lemmas and corollaries (their total must be close to 300)."(Zariski, 1948, 674). Thus, it can be claimed that, although the choice of exposition made by Weil was close to what one finds in Bourbaki—the Satz-Beweis type —, the patterns of definitions and inferences were not. In fact, as we will argue, there is another important aspect of Bourbaki's style that Weil does not quite follow to its natural conclusion in his work.

It is important to note that Bourbaki's volumes are expository. They are not research monographs, even though some of them include some very recent developments at the time of their writing. But I do not believe that the analysis that I propose is limited to these expository works. It can and was adopted by some of Bourbaki's members. I would claim, for instance, that Chevalley and Grothendieck both have produced mathematics that exhibit Bourbaki's style, although in the case of Grothendieck, it is a structuralist style that is a variant or an extension of Bourbaki's. These are empirical claims that will have to be established by looking at their work if my analysis holds any water.

Let us now briefly look at the mode of presentation of the material chosen by Bourbaki. We first want to emphasize one aspect that, although important in the organization and the presentation of the material, does not constitute, in my opinion, an essential aspect of Bourbaki's structuralist style.

8.3.3 Bourbaki's Writings

Every book of Bourbaki's *Éléments de mathématique* comes with a user guide.[15] They all open with a warning "To the Reader". The first paragraph goes like this:

> 1. This series of volumes, [...], takes up mathematics at the beginning, and gives complete proofs. In principle, it requires no particular knowledge of mathematics on the reader's part,

[15] Once again, I do not claim any originality in this section. But it is essential to untangle different components present in the writings.

> but only a certain familiarity with mathematical reasoning and a certain capacity for abstract thought. [...] (Bourbaki, 2004, v)

It is no accident that Bourbaki insists right from the beginning on a 'certain capacity for abstract thought.' We will argue that it is in fact a crucial part of Bourbaki's mathematical style. The next paragraph goes into more detail.

> 2. The method of exposition we have chosen is axiomatic and abstract, and normally proceeds from the general to the particular. This choice has been dictated by the main purpose of the treatise, which is to provide a solid foundation for the whole body of modern mathematics. For this it is indispensable to become familiar with a rather large number of very general ideas and principles. Moreover, the demands of proof impose a rigorously fixed order on the subject matter. It follows that the utility of certain considerations will not be immediately apparent to the reader... (Bourbaki, 2004, v)

Notice that this solid foundation rests on the abstract axiomatic foundation, not explicitly on logic and set theory, although the first volume is indeed on logic and set theory. They certainly play a role and are part of the style, but it is clear that the weight is placed on the abstract axiomatic method which is grounded on them. Logical aspects of the volumes are nonetheless identified immediately. Logic plays two important roles in the enterprise. The first one is global and described in paragraph 4:

> 4. This series is divided into volumes (here called "Books"). The first six Books are numbered and, in general, every statement in the text assumes as known only those results which have already been discussed in the preceding volumes. This rule holds good within each Book, [...]. At the beginning of each of these books (...), the reader will find a precise indication of its logical relationship to the other Books and he will thus be able to satisfy himself of the absence of any vicious circle.

Thus, there is a global logical organization of the whole books. It is systematic and coherent.

The second one is local and shows up in the following paragraph.

> 5. The logical framework of each chapter consists of the *definitions*, the *axioms*, and the *theorems* of the chapter. These are the parts that have mainly to be borne in mind for subsequent use (Bourbaki, 2004, vi).

This is now a specific mode of presentation, based on a logical framework. And indeed, anyone who has looked at and studied mathematics by reading Bourbaki is struck by the following facts, which make it hard not to fall back on Mattuck's adjectives. The presentation can only be qualified as being extremely dry, severe, austere, unified and terse. There are no images, no informal motivations or descriptions, no explanations of the value of this theorem or that definition. But at the same time, it is clean, elegant, and efficient. As some say that there are no unnecessary notes in Mozart's music, there are no unnecessary definitions, axioms, theorems, lemmas and examples in Bourbaki's mathematics. Another comparison readily comes to mind: Bourbaki's organisation of the material is akin to the plans of the architects of the Bauhaus school and their students.

Here is how Cartan described these components in 1958, at the heyday of their production. Not surprisingly, we find the same elements contained in the note to the reader.

> All the books of part I are arranged from a strictly logical point of view. A concept or result may be used only if it has appeared in a previous chapter of a book. Obviously, one has to pay a high price for such rigor: the resulting presentation tends to become somewhat ponderous. The reader finds its weightiness repellant, and the style is certainly not what one would call inspiring. The mathematical text consists of a series of theorems, axioms, lemmas, etc. This rigorous, precise style stands in sharp contrast to the light and not too precise style of the French tradition at the end of the last century. [...] Today it is apparent that this precise style is finding its way more and more into mathematical literature (Cartan, 1979).

Nothing is presupposed. Everything is defined from scratch and thereafter, the proofs all depend on notions and theorems already given and proved. This is an adequate characterization of Bourbaki's expository style. But I argue that it does *not* give us, as such, Bourbaki's mathematical style.

In a different paper, written much earlier, Cartan makes the following remarks about the logical component and the epistemic component of the axiomatic method:

> Now suppose these axioms chosen once and for all. Our mathematical theory must not restrict itself to be a dull compilation of truths, that is of consequences of axioms that we note, for each and every one of them, laboriously the accuracy. For mathematics to be an effective instrument and, also, for us, mathematicians, to be able to take a true interest in it, it must be a living construction: one must clearly see the web of theorems, group the partial theories. In this task, *it is again the axiomatic method that comes to our help*, by giving us the *principle of classification*. [...] Today, more and more we tend to study algebraic structures, topological structures, and ordered structures, etc. [...]
> Thus, not only the axiomatic method, based on pure logic, gives a steadfast seat to our science, but it also allows us to organize it better and to understand it better, it makes it more effective, it substitute general ideas to "computations" that, carried out haphazardly, would most likely lead nowhere, unless done by an exceptional genius. (Cartan, 1943, 11) [my translation and emphasis]

Thus, we have to distinguish the logical dimensions of the axiomatic method from the epistemic dimensions. The epistemic dimensions are built upon the logical ones. A purely logical presentation of mathematics already existed when Bourbaki wrote their books: it was given by Russell and Whitehead *Principia Mathematica*. Granted, it was not based on sets, and in some respects it was a failure, but it certainly was rigorous, austere and precise. Bourbaki and some of its members did publish on the logical foundations of mathematics[16] And Bourbaki did claim that he wanted to derive the whole of mathematics from the axioms of set theory.[17] It is clear that there is a polemical element present in these papers, in particular the first two. Indeed, they present the foundational program in the spirit of Hilbert's answer to Brouwer. It is therefore tempting to reduce Bourbaki's project to its logical development. We believe that this move is, however, far too quick. For one

[16] See Dieudonné (1939), Cartan (1943), and Bourbaki (1949).

[17] This is explicit in Bourbaki (1949).

thing, Bourbaki did not want to include logic in their project at first. And Bourbaki always looked at logic as a mere instrument, as providing the proper grammar of mathematics.

8.3.4 Bourbaki's Style

Before we apply our general definition of mathematical style to Bourbaki, we first have to present and discuss Chevalley's article published in 1935 and entitled "Variations of mathematical style", (Chevalley, 1935), in the *Revue de Métaphysique et de Morale*.[18] Interestingly, while Bourbaki was coming to life, one of its members published a paper in a philosophy journal that discusses precisely the notion of mathematical style.[19]

8.3.4.1 Chevalley on Mathematical Style

As we have already indicated, when discussing Rabouin's analysis of the notion of mathematical style in the foregoing section of our paper, Chevalley does not give a general definition of mathematical style. He identifies three different mathematical styles in his paper: the style based on infinitesimals, the ϵ-style and the axiomatic style. Each one of these is characterized by contrasting it with the preceding style.

Chevalley opens up his paper by saying that he is not interested in the *personal* style of some mathematician, but rather the style of a period, a general tendency that becomes the norm under the influence of certain individuals. To illustrate what he means, he presents the "ϵ-style", a style forged under the influence of Weierstrass.

The ϵ-style itself has a history and became the norm when infinitesimals were seen to lead to difficulties. Thus, the desire to bring rigor into some mathematical demonstrations, in particular those involving infinitesimal quantities, brought about important changes in the practice of mathematics. Not that infinitesimals were not useful; thinking and doing mathematics with infinitesimal quantities was fruitful, even fertile. But their use needed some justification and it opened the door to some anomalies, for instance Weierstrass's discovery of a continuous real function in one

[18] The title is *Variation du style mathématique* in French.

[19] Chevalley is a very interesting case. Not only was he a brilliant mathematician, but he was also interested in politics, philosophy and the foundations of mathematics. As he himself revealed later, he was solely responsible for the inclusion of logic in Bourbaki's books. One of his best friends was Jacques Herbrand, a brilliant young logician who unfortunately died in a hiking accident in 1931 at the age of 23. Albert Lautman, the philosopher of mathematics who was killed by the Nazis in 1944, was also one of his good friends. Later in his life, Chevalley joined Grothendieck and founded the movement *Survivre et vivre* in Montreal during the summer of 1970. It might also be worth mentioning that Chevalley studied under Emil Artin in the early 1930s and then with Helmut Hasse, both vigorous developers of what was then called the "axiomatic method".

variable that is nowhere differentiable, but which can be defined nonetheless by an ordinary looking Fourier development. It appeared that something was wrong somewhere. Similar functions could show up in classical analytic theories without notice. It was by trying to clarify the foundations of these infinitesimal quantities that a new mathematical style emerged. This style, according to Chevalley, can be identified by certain obvious traits.

As its name indicates, the usage, sometimes immoderate according to Chevalley, of various ϵ, with indices, is the most obvious feature of that style. The progressive replacement of equalities by inequalities in proofs, theorems, etc. is the second sign. Notice immediately that the components of the style identified by Chevalley are argumentative strategies, ways of proving that are brought in to make mathematics more rigorous. Although he does not mention what we called 'means of definitions' explicitly, he certainly could have done so.

According to Chevalley, it is precisely this reliance on inequalities that inevitably lead to the limitation of that style and the need to develop a different style.

> Indeed, while equality is a relation that makes sense for arbitrary mathematical beings, inequality can only bear upon objects provided with a certain order, in practice only on real numbers.[20] This therefore leads, in order to embrace the whole of analysis, to the complete reconstruction from real numbers and functions of real numbers. ... One could believe at some point that mathematics would constitute itself in a unitary domain, founded entirely by constructive definitions from the real numbers. (Chevalley, 1935, p. 379) [Our translation]

He simply states that this unification did not happen. For, some mathematical concepts cannot be constructed from the real numbers, for instance the concept of group. Geometry, although it can be constructed to a certain extent in the ϵ style, becomes somewhat ad hoc or artificial. The nature of points, as n-tuples of real numbers, is not essential to geometry, as Klein had conclusively shown. It is the group of transformations of a geometry that provides the equality of figures inherent to that geometry, not the equality of points. Thus, in some cases, constructive definitions provided by analysis *hide* the real nature of what they were trying to define.

Chevalley then states that geometry provided, in fact, the material of what was to become the new style. He attributes the emergence of this way of doing mathematics to Hilbert's work in geometry.[21] One does not construct points, lines, planes, and other geometric objects from more primitive notions, but rather one simply stipulates, by stating axioms, some of their fundamental properties, leaving the nature of the objects completely undetermined. One then proceeds by proving theorems from these axioms and then note that the points of the geometry can be associated to points of real numbers and that the axioms of the theory

[20] It is interesting to see that Chevalley does not consider abstract ordering structures at that point. He did not know about Birkhoff's or Ore's work at the time. See Corry (2004) for more on the latter. Bourbaki will later on think of order structures as fundamental to mathematics.

[21] Whether this is historically adequate, we will simply ignore. It is a debate among Hilbert scholars that need not concern us here.

are true when geometric points and planes are replaced by objects constructed from real numbers.[22] Hilbert's success apparently inspired other mathematicians. Chevalley mentions Lebesgue's integral, which is given by a list of properties and the concept of topological space, in which Weierstass limits are obtained from a purely abstract characterization, as Fréchet has shown. And then, there is algebra. Chevalley points out that one could even claim that the whole movement in fact originates from that source, more precisely from Dedekind's work and teaching of abstract groups. However, in algebra, Chevalley claims that the turning point can be found in Steinitz's work on field theory. Chevalley then claims that "the axiomatization of theories has profoundly changed the style of contemporary mathematical writings"(Chevalley, 1935, p. 381).

Thus, the hallmark of the new style is the axiomatic method. Chevalley already emphasizes the fact that the axioms are not chosen arbitrarily. Mathematicians start from given, known proofs. One then performs an analysis of these proofs and tries to identify the properties that are strictly necessary to obtain a given result. One looks for the minimal logical requirements and tries to identify the domain of mathematics in which the result can be proved. Once this is done, it is possible to eliminate unnecessary hypotheses. In this way, according to Chevalley, one obtains elegant demonstrations. Chevalley, in 1935, identifies the autonomous domains of mathematics: in algebra, field theory, the theory of abstract groups, ring theory, hypercomplex numbers (now known as algebras); in analysis, measure and integration theory, topology, Riemann surfaces, Hilbert spaces; in geometry, projective and conformal geometries, Riemann spaces, combinatorial topology (renamed algebraic topology soon afterwards). In each case, we get a specific type of *abstract structure*. Chevalley then claims that these theories combine, yielding, for instance, topological groups, which is seen as a new *abstract structure*. In other cases, some theories turn out to be based on the same axioms, or, in the words of Chevalley, their axioms yield the same structure, as is the case of probability theory and measure theory.

Traditional mathematical objects emerge from the combinations and interactions of some of these abstract structures. Chevalley mentions the system of real numbers: it is a field, a topological space, a topological group, an ordered set, a measured space, etc. The properties of real numbers are either theorems of one of these abstract structures that apply to them, or "properties resulting from the simultaneous validity of many of these theories"(Chevalley 1935, p. 383[our translation]). It is worth quoting the closing paragraph of Chevalley's paper:

> It results from all this that contemporary mathematics tries to define mathematical objects in comprehension, that is by their characteristic properties, and not by extension, that is by construction. This aspect is undoubtedly not definitive. But it is hard to predict at this point in which direction it will evolve. Be that as it may, the actual tendency seems far from having exhausted its internal dynamism. The diverse theories that have been separated up until now probably have not attained their definitive form. Many of them will probably be analyzed in terms of superpositions of even more general theories; others will turn out to be

[22] In the paper, Chevalley does not talk about interpretations, but by replacing one by another.

> equivalent with one another or deriving from a common source. The structural analysis of facts already known is far from being done, not mentioning the analysis of these new facts that manifest themselves once in a while. (Chevalley 1935, p. 384)[our translation]

Chevalley is thus well aware that these autonomous domains, as he calls them, might change as mathematics evolves.

The last sentence of the paragraph is, for us, revealing: one has to effectuate a *structural analysis* of facts. Both words are important: one looks for a *structure* and it is obtained via an *analysis*. Both words are philosophically loaded and have a long history. Chevalley was certainly aware of that. Be that as it may, the expression captures perfectly the basis of the new style. The structural analysis leads to the identification of one or more structures and the latter are then explicitly captured by the axiomatic method.[23]

In his paper, Chevalley provides a sketch of the 'new point of view' underlying Bourbaki's structuralist standpoint, Bourbaki's style.[24] It is striking to see that the paper *The Architecture of Mathematics* and other papers written in the 1940s by various members of the collective essentially repeat and expand on what Chevalley had already said in 1935. They appear to be nothing more nor less than a more precise and updated version of the same ideas.

8.3.4.2 Bourbaki's Epistemic Mathematical Style

Contemporary mathematicians did not hesitate to talk about Bourbaki's style in such a way that it pointed towards something more than simply the writing style. But no one clearly provided a characterization of the style. Halmos made a direct comparison with music:

> The Bourbaki style and spirit, the qualities that attract friends and repel enemies, are harder to describe. Like the qualities of music, they must be felt rather than understood (Halmos, 1957).

There is no doubt that the style of presentation must be felt, but we will nonetheless propose a characterization of Bourbaki's epistemic style by applying our general framework.

First, we have to identify γ, the background culture, more specifically the standards against which Bourbaki's style has to be compared and contrasted. Since their original goal was to write a textbook on analysis, the background is given by

[23] But it might also be somewhat too short. In his paper *Mathématiques et réalité* published in 1936, Albert Lautman characterizes the work of the Hilbert school as providing 'the synthesis of necessary conditions and not that of the analysis of first notions.'(Lautman, 2006, 49). Lautman is emphasizing the synthetic component inherent to the process of abstraction as embodied in the axiomatic method. He also explicitly refers to Carnap's work and the role of analysis in the latter.

[24] Patras, in his book Patras (2001), takes a similar position with respect to the idea that Bourbaki adopts a certain mathematical style.

the French textbooks of the time, those that they were using themselves.[25] One of the texts used at the time was Émile Goursat's *Cours d'analyse mathématique*.[26] Even a cursory look at Goursat's books indicates that it is an instance of what Chevalley called the ϵ-style, with many epsilons. Needless to say, Goursat does not use the language of sets systematically, the definitions are informal, in the sense that there is no explicit logical apparatus and there are no abstract structures involved either.

Let us move to δ, the patterns of definition. First, Bourbaki decides to use systematically, explicitly and in all cases the language of set theory as expressed in first-order logic. As we have already mentioned, no less than three different papers were written, namely by Jean Dieudonné, Henri Cartan and André Weil, in the late 1930s and 1940s, the latter presented by André Weil at the meeting of the Association of Symbolic Logic, to emphasized the need to provide explicit logical foundations for the working mathematicians. They are clearly not interested in the logical foundations of mathematics for its own sake, nor do they see in the latter as having any real impact on the work of mathematicians.[27] In fact, this choice has to be put in the perspective of the cultural background γ. For no one before Bourbaki had explicitly decided to present concretely in a unified manner, in one language, all the concepts required to do analysis, and, as they thought could be done in the 1930s and 1940s, the whole of mathematics. We will not belabor the idiosyncratic system of axioms of set theory chosen by Bourbaki, for what matters to us is merely the fact that they adopted the language of sets and the formalism of first-order logic in their presentation and practice.[28]

It is in this language that the axiomatic method is used to define abstract structures. But we have to be clear as to what is meant here; it is not merely that a mathematician postulates what she likes and derives theorems from there. The axioms that come at the beginning of a presentation are in fact the result of a "structural analysis", to use Chevalley's words, and they are put together, thus synthesized, into a new, autonomous whole. The same idea appears later. In the

[25] Needless to say, it would be relevant to do more detailed historical research and look carefully at the textbooks that were in circulation in France in the 1920s and early 1930s. We know that the original members of Bourbaki knew about and were influenced by books published outside of France, e.g. van der Waerden's *Moderne Algebra*, Seifert & Threlfall's *Lehrbuch der Topologie*, Alexandroff & Hopf's *Topologie*, Lefschetz's *Topology*, among others. We rely here on Beaulieu (1990), Corry (2004), and Houzel (2004).

[26] Goursat's books can be consulted online at https://archive.org/details/coursdanalysema00gourgoog/mode/2up.

[27] Again, Chevalley is the only one who seemed to have taken a genuine interest in foundational studies at the time. He even wrote a report on Gödel's work on the consistency of the continuum hypothesis and I suspect that Gödel's work did influence him in his thinking as to how to give a general metamathematical account of the notion of structure. But this specific point will have to be argued elsewhere. For his report, see http://sites.mathdoc.fr/archives-bourbaki/PDF/065_iecnr_074.pdf.

[28] For critical evaluations of the axiomatic system adopted by Bourbaki, see Mathias (1992) and Anacona et al. (2014).

1940s, under the name of Bourbaki, Dieudonné wrote:

> Today, we believe, however, that the internal evolution of mathematical science has, in spite of appearance, brought about a closer unity among its different parts, so as to create something like a central nucleus that is more coherent than it has ever been. *The essential aspect of this evolution has been the systematic study of the relations existing between different mathematical theories, and which led to what is generally known as the "axiomatic method"*. (Bourbaki, 1950, 222)[my emphasis]

The function of the axiomatic method is to *abstract* new, original concepts from classical settings, and then to use this to reconstruct and extend these classical results in new directions. The idea is expressed later by Cartan:

> From the beginning, Bourbaki was a decided supporter of the so-called axiomatic method. [...] How does it [the axiomatic method] apply to higher mathematics? A mathematician setting out to construct a proof has in mind well defined mathematical objects which he is investigating at the moment. When he thinks he has found the proof, and begins to test carefully all his conclusions, he realizes that only a very few of the special properties of the objects under consideration have played a role in the proof at all. He thus discovers that he can use the same proof for other objects which have only those properties he had employed previously. Here we can see the simple idea underlying the axiomatic method: instead of declaring which objects are to be investigated, one only has to list those properties of the objects to be used in the investigation. These properties are then brought to the fore expressed by axioms; whereupon it ceases to be important to explain what the objects *are*, that are to be studied. [...] It is quite remarkable how the systematic application of such a simple idea has shaken mathematics so completely (Cartan, 1979, 176–177).

This passage emphasizes the standards γ of the time again: when Bourbaki started to work on their project, this so-called axiomatic method was not systematically used in this way. There were important examples of its use in diverse areas, but it was not conceived as a way to reconstruct the whole of mathematics, as a way to introduce mathematical structures in general.

We are clearly dealing with a special type of axiomatic method which is now part of a new set of patterns of definition. The axioms are merely a contingent vehicle to talk about the concept of an *abstract* mathematical *structure*. The first step of the axiomatic method is to excavate the essential working components in diverse mathematical situations and extract or abstract the properties, operations, relations, etc. that are then expressed in the axioms. The latter provide a structure, an object of study in itself. Structures are related to one another in ways that classical mathematical fields were not, that is, by the properties, operations, relations that are abstracted out. It thus leads to a complete reorganization of mathematics and a completely different understanding of mathematical concepts.

> Bourbaki's decision to use the axiomatic method throughout brought with it the necessity of a new arrangement of mathematics' various branches. It proved impossible to retain the classical division into analysis, differential calculus, geometry, algebra, number theory, etc. Its place was taken by the concept of *structure*, which allowed definition of the concept of isomorphism and with it the classification of the fundamental disciplines within mathematics (Cartan, 1979, 177).

This last sentence by Cartan captures an essential part of Bourbaki's epistemic style: "the concept of structure... allowed definition of the concept of isomorphism

and with it the classification of the fundamental disciplines within mathematics." Thus, Bourbaki's patterns of definition of structures include intrinsically the notion of isomorphism. The latter is built in, it is part of the axioms, thus the definitions and, it will be part of the inference patterns, as we will see. Alas, Bourbaki's formal characterization of the notion of mathematical structure is often seen as a failure. We strongly believe that to discard it completely is a mistake; there is no need to throw the baby out with the bathwater.

8.3.4.3 Bourbaki's Definition of Abstract Mathematical Structures and Isomorphisms

The importance of including the notion of isomorphism in the very definition of structures was understood early by Bourbaki. Here is how Chevalley expressed it in an unpublished version of the introductory chapter on sets:

> There are finally cases where the content of thought refers almost uniquely to the formal aspect of the notion considered. This is how, when a mathematician thinks of the content of the idea that he has of isomorphic mathematical beings, he will note, we believe, that he thinks less of the complete similarity of two objects as things than the following: any theorem concerning one of these objects can be translated into a theorem concerning the other. Chevalley, 26 [my translation]

Chevalley expresses a very important shift in this quote, a shift that will be included in the final version of Bourbaki's technical definition of species of structure. We move from the idea of isomorphic mathematical beings in terms of similar objects to the claim that they are objects that satisfy the same theorems of a theory, or, from a proof-theoretical point of view, that the same theorems can be proved about these isomorphic beings. Thus, the idea is to define structures with the notion of isomorphism built in, so that if a specific theorem about one of these structures is proved, then any structure isomorphic to it will satisfy the very same theorem. Moreover, the only theorems such a theory ought to be able to prove are precisely those that are invariant under isomorphism. Thus, the pattern of definition includes a pattern of inference. This is the key component of the structuralist style.

Bourbaki's published technical definition of a "species of structure" is indisputably clumsy and was recognized as such. Moreover, and as we will briefly indicate later, when the final version was finally accepted by Bourbaki, they were very well aware that their definition could not accommodate categories and functors, and after many different attempts by different members, even Eilenberg, one of the creators of category theory, they simply gave up and published their latest attempt, which could only cover set-based structures.

I will not sketch Bourbaki's technical definition. I will rather offer a reconstruction of Bourbaki's notion of species of structure.[29] There are two reasons for presenting the reconstruction rather than Bourbaki's published version. First,

[29] We thank Michael Makkai for this reconstruction.

we will use a more standard and transparent presentation. Second, it will be clear that Bourbaki's definition, which is really a different way of introducing the same ideas, is fully metamathematical. Indeed, in their final published version, when the reader finally gets to the definition of a species of structure, he or she reads "*A species of structures* in $\mathfrak{T}$ is a text Σ formed of..."(Bourbaki, 2004, 262). Look at it again: a species of structures is a *text*. How should one interpret this sentence? Is Bourbaki adopting a formalist stance? Notice that it is consistent with Chevalley's position with respect to mathematical style: it is a way of writing. It is nonetheless clear that Bourbaki's formal set-up has a natural interpretation in a universe of sets. It is a text with a canonical interpretation. More specifically, a species of structures has to be given by a formulas in a language, and when interpreted, it is a set together with relations, etc. But what this clearly indicates is that Bourbaki is firmly, when he writes this, in *metamathematics* and not in mathematics.[30] This is methodologically very important, for it translates concretely the idea contained in Chevalley's foregoing quote. It is only in a metamathematical framework that one can state in full generality the requirement that isomorphic structures satisfy the same theorems. Moreover, one needs a fully general notion of isomorphism, something that did not exist when Bourbaki started to define species of structures, and this point has to be taken as an additional deviation from γ. Now, to the reconstruction.

We work in first order logic. Let $\vec{X} = X_1, \ldots, X_n$, a finite list of *basic set variables*[31] and $\vec{B} = B_1, \ldots, B_m$ another sequence of *parameters*. The latter are necessary to cover cases like vector spaces over a field k, modules over a ring R, etc.

Definition 1 An echelon construction on the set variables $X_1, \ldots, X_n$ and parameters $B_1, \ldots, B_m$, is a collection S of terms defined inductively as follows:

1. Each of $X_1, \ldots, X_n, B_1, \ldots, B_m$ is in S;
2. If A_1 and A_2 are in S, so is $A_1 \times A_2$;
3. If A is in S, so is $\mathcal{P}(A)$.[32]

This is a standard inductive definition which gives us terms, that is denoting expressions, constructed in a systematic fashion.

Thus, an echelon construction S gives us the basic terms that are given or have to be constructed for the structure of a given kind to be defined. Let us denote an

[30] Granted, there is a clear shift in the section on structures. Bourbaki undisputably starts in a metamathematical framework, but as the section develops and tries to incorporate concepts that clearly belong to category theory, it morphs into a mathematical mode. It is a case of conceptual schizophrenia.

[31] We follow Bourbaki for the time being and talk about sets. They really are simply formal variables that will stand for sets. As variables, they are distinct.

[32] Some readers might be struck by the fact that we seem to be moving towards a type theory. Indeed, in many early versions of the theory of species of structures, Bourbaki does work with types. He progressively abandons the type theoretical terminology in favor of a purely set theoretical.

element of an echelon construction S by s_i and we will call such an element a *sort*. We can now introduce the notions of a similarity type, which was not in Bourbaki but is standard in logic.

Let S be an echelon construction and $\overrightarrow{S} = s_1, \ldots, s_p$ a sequence of chosen elements of S. These are now called *specified sorts*.

Definition 2 A signature $\mathcal{L} = \mathcal{L}(\overrightarrow{X}, \overrightarrow{B}, \overrightarrow{S}, \overrightarrow{R})$ (or similarity type) is given by:

1. A list $\overrightarrow{X} = X_1, \ldots, X_n$ of (basic set-)variables;
2. A list $\overrightarrow{B} = B_1, \ldots, B_m$ of parameters;
3. A list of specified sorts $\overrightarrow{S} = s_1, \ldots, s_p$, each $s_i \in S$;
4. A list of relation symbols $\overrightarrow{R} = R_1, \ldots, R_p$, each R_j specified as a (sorted) relational symbol $R_j \subset s_j$, more precisely the arity of R_j is $R_j \subset s_{i_{j,1}} \times s_{i_{j,2}} \times \cdots \times s_{i_{j,k_j}}$.[33]

This is all purely formal. We are just setting up the syntactic framework that allows us to talk about structures. In fact, we are now in a position to specify what a structure for a given signature $\mathcal{L}(\overrightarrow{X}, \overrightarrow{B}, \overrightarrow{S}, \overrightarrow{R})$ is.

Definition 3 An $\mathcal{L}$-structure M is given by the following data:

1. A tuple $\overrightarrow{X}^M = X_1^M, \ldots, X_n^M$ of (not necessarily distinct) sets, the basic sets;
2. A tuple $\overrightarrow{B}^M = B_1^M, \ldots, B_m^M$ of sets, the parameter sets;
3. A tuple $\overrightarrow{S}^M = s_1^M, \ldots, s_p^M$ of derived sets; each of these is understood as the set-interpretation of the corresponding echelon term; with the given sets $X_1^M, \ldots, X_n^M, B_1^M, \ldots, B_m^M$ plugged in for the variables $X_1, \ldots, X_n, B_1, \ldots, B_m$ respectively;
4. Actual relations $R_1^M, \ldots, R_p^M$ with R_j^M a relation of the type

$$R_j^M \subset s_{i_{j,1}}^M \times s_{i_{j,2}}^M \times \cdots \times s_{i_{j,k_j}}^M;$$

$$R_j^M \subset s_j^M.$$

Now, the parameter sets, although arbitrary are fixed for a given structure. We will make that explicit in the notation.

Let us fix $\overrightarrow{B} = \underline{B}_1, \ldots, \underline{B}_m$, the parameter sets. Notice the change in the notation here: we denote an actual, fixed set by an underline $\underline{B}$. We define an $\mathcal{L}_{\overrightarrow{\underline{B}}}$-*structure* to be an $\mathcal{L}(\overrightarrow{X}, \overrightarrow{B}, \overrightarrow{S}, \overrightarrow{R})$-structure M where $B_i^M = \underline{B}_i$ for $1 \leq i \leq n$.

So far, we haven't done anything extraordinary or difficult. We have given a simple type of $\mathcal{L}$-signature and $\mathcal{L}$-structure. The only original element comes from the echelon construction underlying both definitions. We hasten to add that this

[33] Needless to say, functions can be introduced as special kind of relations, as usual.

notion of $\mathcal{L}$-structure is *not* (yet) the notion we are driving at. We still have to impose a restriction on the latter to get to the notion of a *Bourbaki species of structure*. But for that, we need to define isomorphism and isomorphism transfer for $\mathcal{L}_{\underline{\vec{B}}}$-structures.

Isomorphism and Transport of Structure

We start with two n-tuples of basic sets $\overrightarrow{X}^1 = X_1^1, \ldots, X_n^1$ and $\overrightarrow{X}^2 = X_1^2, \ldots, X_n^2$. We assume we are given an echelon construction S and an element s of S. We now fix the following notation. The *interpretations* of $s_{\overrightarrow{X}^1}$ and $s_{\overrightarrow{X}^2}$ of s is given inductively as follows:

1. If s is X_i, then $s_{\overrightarrow{X}^1}$ is X_i^1 and $s_{\overrightarrow{X}^2}$ is X_i^2;
2. If $s = B_i$, then $s_{\overrightarrow{X}^j} = \underline{B}_i$ for both $j = 1, 2$;
3. If $s = s_1 \times s_2$, then $s_{\overrightarrow{X}^j} = (s_1)_{\overrightarrow{X}^j} \times (s_2)_{\overrightarrow{X}^j}$ for $j = 1, 2$;
4. If $s = \mathcal{P}(s')$, then $s_{\overrightarrow{X}^j} = \mathcal{P}((s')_{\overrightarrow{X}^j})$ for $j = 1, 2$.

The foregoing is straightforward bookkeeping and is merely an exercice in notation and substitution.

Assume that we are given a tuple $\overrightarrow{\phi} = \langle \phi_1, \ldots, \phi_n \rangle$ of bijections $\phi_i : X_i^1 \to X_i^2$, for $i = 1, \ldots, n$. The parameters B_j's are not part of the bijection tuple.

The bijection-tuple $\overrightarrow{\phi}$ induces bijections, for every s in S

$$\phi_s : s_{\overrightarrow{X}^1} \to s_{\overrightarrow{X}^2},$$

in the obvious way, where we use the identity maps $1_{\underline{B}_i} : \underline{B}_i \to \underline{B}_i$.

We can now explain how to transfer an $\mathcal{L}$-structure M to an $\mathcal{L}$-structure N.

Let M be an $\mathcal{L}_{\underline{\vec{B}}}$-structure with the basic sets $\overrightarrow{X}^1 = X_1^1, \ldots, X_n^1$ interpreted as $X_1^M, \ldots, X_n^M$, respectively and, similarly, let N be an $\mathcal{L}_{\underline{\vec{B}}}$-structure with the basic sets $\overrightarrow{X}^2 = X_1^2, \ldots, X_n^2$ interpreted as $X_1^N, \ldots, X_n^N$, respectively. We use the bijections $\phi_i : X_i^M \to X_i^N$ to transfer the $\mathcal{L}_{\underline{\vec{B}}}$-structure M to the $\mathcal{L}_{\underline{\vec{B}}}$-structure N as follows.[34]

Definition 4

1. For each of the sorts $s_1, \ldots, s_p$,

$$s_j^N = (s_j)_{\overrightarrow{X}^2};$$

[34] These are Bourbaki's 'transportable relations'.

2. For each of the relation symbols $R_1, \ldots, R_p$, $R_j \subset s_j$ with arity $R_j \subset s_{i_{j,1}} \times s_{i_{j,2}} \times \cdots \times s_{i_{j,k_j}}$, we have the interpretation

$$R_j^M \subset s_{i_{j,1}}^M \times s_{i_{j,2}}^M \times \cdots \times s_{i_{j,k_j}}^M,$$

$$R_j^M \subset s_j^M$$

together with the bijective mappings

$$\phi_{s_{i_{j,1}}} \times \cdots \times \phi_{s_{i_{j,k_j}}} : s_{i_{j,1}}^{\overrightarrow{X}^1} \times \cdots \times s_{i_{j,k_j}}^{\overrightarrow{X}^1} \to s_{i_{j,1}}^{\overrightarrow{X}^2} \times \cdots \times s_{i_{j,k_j}}^{\overrightarrow{X}^2}$$

$$\phi_{s_j} : (s_j)_{\overrightarrow{X}^1} \to (s_i)_{\overrightarrow{X}^2}.$$

We define R_j^N as the image of $R_j^M \subset s_{i_{j,1}}^M \times s_{i_{j,2}}^M \times \cdots \times s_{i_{j,k_j}}^M$, $R_j^M \subset s_j^M$ under the foregoing bijective mapping. Thus, R_j^N necessarily satisfies

$$R_j^N \subset s_j^N = (s_j)_{\overrightarrow{X}^2}; \tag{8.1}$$

$$R_j^N \subset (s_{i_{j,1}})^N \times \cdots \times (s_{i_{j,k_j}})^N, \text{ that is} \tag{8.2}$$

$$R_j^N \subset s_{i_{j,1}}^{\overrightarrow{X}^2} \times \cdots \times s_{i_{j,k_j}}^{\overrightarrow{X}^2}. \tag{8.3}$$

This definition can be captured by the following diagram:

$$\begin{array}{ccc} (s_j)_{\overrightarrow{X}^1} & \xrightarrow{\phi_{s_j}} & (s_j)_{\overrightarrow{X}^2} \\ \uparrow & & \uparrow \\ R_j^M & \dashrightarrow^{\simeq} & R_j^N \end{array}$$

where the dotted arrow signifies that it is induced by the given data to make the diagram commute.

We thus obtain an *isomorphism* $\widehat{\overrightarrow{\phi}} : M \xrightarrow{\simeq} N$ that is completely determined by the given bijection-tuple $\overrightarrow{\phi}$ on the basic sets $\phi_i : X_i^M \xrightarrow{\simeq} X_i^N$, $i = 1, \ldots, n$, that *preserves* the relations $R_1, \ldots, R_p$.

Now let us be given a set-theoretic formula

$$\Phi(\overrightarrow{X}, \overrightarrow{R}, \overrightarrow{B})$$

with the same free variables as before and no more. We *assume* that the formula Φ implies, that is contains as conjuncts, the specifications that

$$R_j \subset s_j \tag{8.4}$$

$$R_j \subset s_{i_{j,1}} \times s_{i_{j,2}} \times \cdots \times s_{i_{j,k_j}}. \tag{8.5}$$

We have a standard formula $\mathrm{Iso}(\overrightarrow{\phi}; \overrightarrow{X}^1, \overrightarrow{R}^1; \overrightarrow{X}^2, \overrightarrow{R}^2; \overrightarrow{B})$ with the distinct free variables as shown that expresses that $\overrightarrow{\phi}$ *is* an isomorphism of the $\mathcal{L}_{\overrightarrow{B}}$-structures M and N

$$\vec{\phi} : M \xrightarrow{\simeq} N$$

where M is given by $\overrightarrow{X}^1$ and $\overrightarrow{R}^1$ and N by $\overrightarrow{X}^2$ and $\overrightarrow{R}^2$. We can now formulate *Bourbaki's condition of isomorphism invariance*:

In the adopted set-theory, it is *provable* that

$$\vdash \Phi(\overrightarrow{X}^1, \overrightarrow{R}^1, \overrightarrow{B}^1) \wedge \mathrm{Iso}(\overrightarrow{\phi}; \overrightarrow{X}^1, \overrightarrow{R}^1; \overrightarrow{X}^2, \overrightarrow{R}^2; \overrightarrow{B}) \Longrightarrow \Phi(\overrightarrow{X}^2, \overrightarrow{R}^2, \overrightarrow{B}^2).$$

This is, of course, the key component of the whole construction and will be part of the notion of species of structures.

Bourbaki's definition of species of structures is now at hand.

Definition 5 A *Bourbaki species of structures* is given by the $\mathcal{L}_{\overrightarrow{B}}$-structures whose relations satisfy the condition of isomorphism invariance.

The crucial element to notice is that the notion of isomorphism is *systematically built into* the *definition* of species of structures. It is defined for all $\mathcal{L}_{\overrightarrow{B}}$-structures before the structure is required to satisfy any condition, any axiom. This is now a norm for all concepts defined with the axiomatic method: one has to make sure that the concept is invariant under the proper notion of isomorphism on the technical sense given above.

We now have one of the main components we were after. Bourbaki's metamathematical analysis of the notion of abstract structure automatically yields a crucial component of Bourbaki's mathematical style. Bourbaki is using the axiomatic method as a mode of definition, but he adds an essential ingredient to it, namely the condition of isomorphism invariance. This is not part of the standard axiomatic method. It was not intrinsic to Hilbert's axiomatic method, nor was it clear that it ought to be built into *all* of mathematics. The notion of abstract structure *comes with* the notion of invariance under isomorphism. Again, this is an important deviation from the standards γ.

This has a direct impact on the patterns of inference ι that are part of Bourbaki's style. Of course, as we have noted, the presentation style is of the form *Satz-Beweis* throughout. The logical structure of the proofs and the logical organisation of the volumes are all explicit. This is all well and good, and indeed is a part of ι. But

there is more, and this additional element has to do with the specifically structuralist component of the style. Although we are in a set-based universe, the species of structures possess all and only the properties they have *as structures*. The patterns of reasonings are therefore constrained to these and only to these. One could therefore say that the reasonings are, in fact, *structure*-based. All the steps, *all* the reasonings have to be done up to isomorphism.

There is an additional aspect to the style that follows from the analysis-synthesis method, i.e. the abstract axiomatic method, and this is the use of a certain type of maximality principle. When one analyses a proof and determines the necessary and sufficient components to get the proof, one thus synthesizes the most abstract structure in which the proof is obtainable—relative to the given language and context. One is therefore naturally lead to axiomatize the most general abstract concept. This is an additional epistemic feature of Bourbaki's style, at least from 1935 until the late 1940s, and that has to be included in the δ. The patterns of definition have a direct impact on the patterns of inference and the interactions κ between δ and ι. Thus, Bourbaki introduces, in Bourbaki (1950), the so-called "mother-structures" and their combinations. The specific mother-structures—algebraic, topological, order —, although perhaps intriguing and thought provoking, are epistemically speaking, only secondary. It is the reasoning modes that matter here and it is these that explain the organization of mathematics that emerges from the structuralist standpoint.

The organization of the first four chapters of Bourbaki's *General Topology* illustrates Bourbaki's epistemic style. Chapter One deals with the structure of topological spaces. Filters and ultrafilters are used to deal with the notion of convergence. These two latter notions are purely structural and are not defined with respect to certain numbers and their properties. In the second chapter, the notion of uniform structure is defined and basically replaces the notion of metric space. Chapter Three moves to topological groups, the generic example of a genuinely new structure emerging from the interaction of two abstract structures, and the notion of uniform structure plays an important role in the presentation. We then move to topological rings and their completions. Once these structures and their properties have been studied, Bourbaki finally introduces the real numbers as a topological group which is the completion of the additive group of rationals. They then extend the field structure of the rationals to the reals. Thus, the real line is a combination of a topological, an algebraic and an order structure.

We are now in position to say more specifically how Weil's way of doing mathematics diverged from Bourbaki's style. As we have seen in the foregoing section, Weil did not always start from the most general abstract structure and move down to the more specific context he was interested in. Indeed, instead of adopting a maximality principle with respect to abstract structures, one could claim that Weil adopted a minimality principle instead. Indeed, as Zariski had noticed in his review, Weil restricted himself to the 'simplest facts about abstract fields and their extensions and the bare rudiments of the theory of ideals'. In contrast, Bourbaki uses modern algebra and abstract structures in general 'to the fullest possible extent'. Furthermore, in his approach to the foundations of algebraic geometry,

Weil did not take into account the idea of working with structures that are invariant under isomorphism. In fact, Weil was recalcitrant towards the idea of automatically attaching a type of morphism to a species of structures. Indeed, in his Corry (1996, 380), Corry quotes a letter from Weil to Chevalley:

> As you know, my honorable colleague Mac Lane claims that every notion of structure necessarily implies a notion of homomorphism, which consists in indicating for each data constituting the structure, those which behave covariantly and those which behave contravariantly [...] What do you think can be gained from this kind of considerations?

Weil, interestedly, was also opposed to categories in general and, perhaps, just for this reason.

Categories and Species of Structures

Of course, Bourbaki species of structures are based on sets even though a species of structures does not automatically come with a set-theoretic notion of morphism. Indeed, Bourbaki explicitly rejects this possibility in the final version of the chapter on structures: "A given species of structures therefore *does not imply* a well-defined notion of morphisms."(Bourbaki, 2004, 272). Bourbaki did not find a way to incorporate categories in their definitions and, with hindsight, many members came to the conclusion that Bourbaki's analysis came short.[35]

Of course, one of the main problems was that some categories cannot be sets. And if one allows for the existence of classes, then problematically there are some operations on categories, e.g. functor categories and functors between those, that are not legitimate. But there is more, and it is important to understand this point. When Bourbaki was thinking about these problems, category theory had not attained its full maturity. In particular, the proper notion of isomorphism for categories had still not been identified properly. Indeed, it appeared in press for the first time in Grothendieck's (1957), and even then it was not properly defined. Thus, to cover categories properly, and in particular, to cover categories in a structuralist fashion, in Bourbaki's style, required a change in the metamathematical analysis and the metamathematical framework. The fact is, categories are more abstract than set-based structures and in that framework, category-based structures have to be defined up to equivalence, not up to isomorphism. Two frameworks deal explicitly with these levels of abstraction and so can be readily employed for reconstructing the structuralist style, namely Makkai's FOLDS, as in Makkai (1998), and Homotopy type theory with the univalent axiom, as in Collective (2013).

Thus, we claim that even for categories Bourbaki's structuralist style is entirely clear and legitimate. The main components of Bourbaki's style are a direct

[35] For instance, see Dieudonné (1970) and Cartier (1998) for their evaluation of the situation. See Corry (1996, 2004) and Krömer (2006, 2007) for a more general analysis. Unfortunately, we still don't have access to all the documents of that period which would allow us to better understand how and why Bourbaki failed to include categories in their enterprise.

consequence of their metamathematical analysis of abstract mathematical structures and, in a sense, the style provides a set of *norms* that guide mathematicians both globally, with the overall organization of mathematics, and locally, with the patterns of definition and patterns of proof.

8.3.4.4 Doing Mathematics Up to Isomorphism: Bourbaki's Legacy

Nowadays, pure mathematics *is* done up to isomorphism. Bourbaki's style has become the norm, the standard. It is not questioned. It is a new norm. Students of pure mathematics are taught mathematics that way. We simply do not explicitly see it as their method;[36] we do not have to. The previous styles could have prevailed; likewise, the Bourbaki style could disappear.[37]

Every field is based on a structure or a combination of structures. Theorems are proved by establishing properties of structures and relations between structures. One gets to classical results by combining and specifying various structures. The whole organization of mathematics is turned upside down. The whole ontology of mathematics is revised.[38] Numbers, geometric figures, etc., are now elements of structures, more or less abstract. The (conceptual) foundations of mathematics—in contrast with the logical foundations of mathematics—are now made up of monoids, groups, rings, modules, fields, vector spaces, topological spaces, measure spaces, partial orders, etc. By specifying properties of those, one gets more structures, and their combinations give rise to genuinely new structures.

When one does mathematics *à la* Bourbaki: one identifies the appropriate *abstract* structures involved in a given context; one looks at the theorems about these structures that are relevant to the given problem; one applies these theorems appropriately and solves the problems by using all and only the abstract properties needed. It is highly abstract. It is elegant. It is clean. It is rigorous. But it is awfully hard, for one needs to learn all about these abstract structures and know how and when to use them. Sometimes, it seems unnecessary, uselessly abstract. Does one *need* to use a theorem about locally compact abelian groups to prove Plancherel's theorem about Fourier transforms of certain functions on the real line? Of course

[36] The status of the alleged proof of the ABC conjecture by Mochizuki rests on a subtle discussion regarding isomorphisms and identities, the abstract and the concrete! See http://www.kurims.kyoto-u.ac.jp/~motizuki/SS2018-08.pdf, Sect. 8.2.

[37] It is certainly evolving. Grothendieck and his school have contributed to this change. The structuralist style nowadays includes categories, functors and working up to equivalence of categories. The introduction of higher dimensional categories makes the style even more abstract than it was.

[38] We use the term 'ontology' in its traditional philosophical sense. We could also use it in its modern, engineering sense of classificatory principle. It is quite interesting to see the evolution of the organization of the field and compare, say how mathematical disciplines were organized around 1920 and in the 1960s. In the sense of a classification of disciplines, the mathematical ontology has radically changed with Bourbaki's work.

not. Plancherel certainly did not prove his theorem by using the structure of locally compact abelian groups. Proceeding that way is sometimes seen as a form of intellectual terrorism or a form of elitism. To some, it is repelling. But it *can* be done this way and there are cognitive benefits to doing so.

8.4 The Structuralist Style

Bourbaki's style is an instance of what might be dubbed the 'structuralist style.' Using our definition of mathematical style, we submit that the structuralist style is based on these interrelated components.

1. Patterns of definition for abstract structures. Bourbaki naturally used the axiomatic method. He was well aware that the term 'axiom' does not refer to its usual epistemological sense. One merely needs a systematic procedure to list properties, relations and how they are connected with one another. Sketches could be used and in the context of higher dimensional categories, might very well be used. We assume that the structures are abstract simply because they have been abstracted from previously given mathematical contexts.[39]
2. These patterns of definition have to include criteria of identity for the abstract structures. Bourbaki helped clarify the general notion of isomorphism for species of structures. At the time, it was the natural criterion of identity to define and use. Nowadays, we know that we need homotopy equivalence, categorical equivalence and higher dimension equivalences. Mathematics is done up to a certain type of isomorphism, the latter being derived from the abstract structures one is working with.
3. An appropriate logical framework is needed to codify the inference patterns inherent in these abstract structures. In a sense, first-order logic was designed specifically to tackle set-based abstract structures. First-order logic, set theory and Bourbaki's structuralism co-evolved from the 1910s until the 1950s. It allowed Bourbaki not only to specify what a structure was, but more importantly what is meant to do mathematics 'up to isomorphism'. Bourbaki required that the properties and relations used to define structures be 'transportable', which is to say that they are invariant under isomorphism. More precisely, Bourbaki required that any property P (and relation) present in the axioms of a species of structure $\mathcal{S}$, satisfy the following structuralist principle: for all X of type $\mathcal{S}$, if $P(X)$ and $X \simeq Y$, then $P(Y)$, where the relation $X \simeq Y$ is the appropriate notion of isomorphism for this species of structure. Nowadays, depending on one's needs and goal, one could use Makkai's FOLDS or homotopy type theory. The main point is that these logical frameworks also satisfy the structuralist principle.

[39] Thus, in this sense, being abstract is a relative property and is not opposed absolutely to being concrete.

These two might also be the first steps towards a different system that still has yet to be defined, but that would be designed as to satisfy the structuralist principle.

4. A systematic framework to combine and compare these abstract structures. Again, the axiomatic method *together with* the notion of isomorphism played that part in Bourbaki's case. It quickly turned out to be inadequate, for the language of categories and functors was more effective and systematic, although more abstract.

Many philosophers of mathematics have claimed that Bourbaki's structuralism had nothing to do with what the philosophers call 'mathematical structuralism.' We have, in a companion paper (Marquis, 2020), argued that those philosophers have misunderstood Bourbaki's structuralism. Bourbaki is unfortunately responsible in part for this state of affairs. We will not rehearse our arguments here. One of the reasons given is that Bourbaki's technical notion of (species of) structure was basically flawed and so was mathematically useless. We disagree with this evaluation, our current claim is clear: Bourbaki exemplifies a mathematical style. And anyone interested in the epistemology of mathematical practice should pay attention to its implications for how we reason in mathematics.

References

Anacona, M., L.C. Arboleda, and F.J. Pérez-Fernández. 2014. On Bourbaki's axiomatic system for set theory. *Synthese* 191(17): 4069–4098.

Beaulieu, L. 1990. *Bourbaki: une histoire du groupe de mathématiciens français et de ses travaux (1934–1944)*. Ph. D. thesis, Université de Montréal.

Beaulieu, L. 1994. Dispelling a myth: questions and answers about Bourbaki's early work, 1934–1944. In *The intersection of history and mathematics*, volume 15 of *Science networks historical studies*, 241–252. Basel: Birkhäuser.

Bourbaki, N. 1949. Foundations of mathematics for the working mathematician. *Journal of Symbolic Logic* 14: 1–8.

Bourbaki, N. 1950. The architecture of mathematics. *American Mathematical Monthly* 57: 221–232.

Bourbaki, N. 2004. *Theory of sets*. Elements of Mathematics (Berlin). Berlin: Springer. Reprint of the 1968 English translation [Hermann, Paris; MR0237342].

Buchholtz, U., F. van Doorn, and E. Rijke. 2018. Higher groups in homotopy type theory. In *LICS '18—33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, 10. New York: ACM.

Cartan, H. 1943. Sur le fondement logique des mathématiques. *Revue scientifique* LXXXI: 3–11.

Cartan, H. 1979. Nicolas Bourbaki and contemporary mathematics. *Mathematical Intelligencer* 2(4): 175–180.

Cartier, P. 1998. Le structuralisme en mathématiques: mythes ou réalité? Technical report, Bures-sur-Yvette.

Chevalley, C. 1950. Livre 1. Théorie des ensembles. Introduction (Chevalley). http://archives-bourbaki.ahp-numerique.fr/items/show/475#?c=0&m=0&s=0&cv=0.

Chevalley, C. 1935. Variations du style mathématique. *Revue de Métaphysique et de Morale* 42(3): 375–384.

Chevalley, C. 1956. *Fundamental concepts of algebra*. New York: Academic Press Inc.

Collective. 2013. *Homotopy type theory—univalent foundations of mathematics*. The Univalent Foundations Program, Princeton; Princeton: Institute for Advanced Study (IAS).

Corry, L. 1992. Nicolas Bourbaki and the concept of mathematical structure. *Synthese* 92(3): 315–348.

Corry, L. 1996. *Modern algebra and the rise of mathematical structures*, volume 17 of *Science networks. Historical studies*. Basel: Birkhäuser Verlag.

Corry, L. 2001. Mathematical structures from Hilbert to Bourbaki: The evolution of an image of mathematics. In A. Bottazzini, U. & Dahan Dalmedico (Ed.), *Changing images in mathematics: From the French revolution to the new millenium*, Studies in the history of science, technology and medicine, 167–186. New York: Routledge.

Corry, L. 2004. *Modern algebra and the rise of mathematical structures*, 2nd ed. Basel: Birkhäuser Verlag.

Corry, L. 2009. Writing the ultimate mathematical textbook: Nicolas Bourbaki's éléments de mathématique. In *The Oxford handbook of the history of mathematics*, ed. E. Robson, and J. Stedall, 565–588. Oxford: Oxford University Press.

Dieudonné, J.A. 1939. Les méthodes axiomatiques modernes et les fondements des mathématiques. *Revue Scientifique* LXXVII: 224–232.

Dieudonné, J.A. 1970. The work of Nicholas Bourbaki. *The American Mathematical Monthly* 77(2): 134.

Granger, G.G. 1968. *Essai d'une philosophie du style*. Philosophies pour l'âge de la science. Paris: Colin.

Grothendieck, A. 1957. Sur quelques points d'algèbre homologique. *Tôhoku Math. J. (2)* 9: 119–221.

Guedj, D. 1985. Nicholas Bourbaki, collective mathematician: an interview with Claude Chevalley. *The Mathematical Intelligencer* 7(2): 18–22.

Halmos, P.R. 1957. Nicolas Bourbaki. *Scientific American* 196(5): 88–102.

Houzel, C. 2004. Le rôle de Bourbaki dans les mathématiques du vingtième siècle. *Gazette des mathématiciens* 100: 53–63.

Jackson, A. 1999. Interview with Henri Cartan. *Notices of the AMS* 46(7): 782–788.

Krömer, R. 2006. La "machine de Grothendieck" se fonde-t-elle seulement sur des vocables métamathématiques? Bourbaki et les catégories au cours des années cinquante. *Revue d'histoire des mathématiques* 12: 119–162.

Krömer, R. 2007. *Tool and object*, volume 32 of *Science networks. Historical studies*. Basel: Birkhäuser Verlag. A history and philosophy of category theory.

Kuhn, T. 1970. *The structure of scientific revolutions*, 2nd ed. Chicago: University of Chicago Press.

Kvasz, L. 2008. *Patterns of change*, volume 36 of *Science networks. Historical studies*. Basel: Birkhäuser Verlag.

Lautman, A. 2006. Mathématiques et réalité. In *Les mathématiques, les idées et le réel physique*, 47–50. Paris: J. Vrin.

Licata, D.R., and E. Finster. 2014. Eilenberg-MacLane spaces in homotopy type theory. In *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pp. Article No. 66, 10. New York: ACM.

Makkai, M. 1998. Towards a categorical foundation of mathematics. In *Logic Colloquium '95 (Haifa)*, volume 11 of *Lecture notes in logic*, ed. J. Makowsky and E. Ravve, 153–190. Berlin: Springer.

Mancosu, P. 2017. Mathematical style. In *The Stanford encyclopedia of philosophy*, ed. E. N. Zalta, (Fall 2017 ed.). Metaphysics Research Lab, Stanford University.

Manders, K. 2008. The Euclidean diagram. In *The philosophy of mathematical practice*, ed. P. Mancosu, 80–133. Oxford: Oxford University Press.

Marquis, J.-P. 2020. Forms of structuralism: Bourbaki and the philosophers. In *Structures Mères, Semantics, Mathematics, and Cognitive Science*, ed. A.P.S. Zipoli. Springer.

Mashaal, M. 2000. *Bourbaki : une société secrète de mathématiciens*. Pour la science. Les génies de la science no 2. Paris: Pour la science.

Mathias, A.R.D. 1992. The ignorance of Bourbaki. *Mathematical Intelligencer* 14(3): 4–13.

Mattuck, A. 1957. Review: Claude Chevalley, Fundamental concepts of algebra. *Bulletin of the American Mathematical Society* 63(6): 412–417.

Patras, F. 2001. *La pensée mathématique contemporaine*. PUF.

Rabouin, D. 2017. Styles in mathematical practice. In *Cultures without culturalism*, 196–223. Durham: Duke Univniversity Press.

Rijke, E., and B. Spitters. 2015. Sets in homotopy type theory. *Mathematical Structures in Computer Science* 25(5): 1172–1202.

Zariski, O. 1948. Review: André Weil, foundations of algebraic geometry, ams, New York, 1946. *Bulletin of the American Mathematical Society* 54(7): 671–675.

Chapter 9
Grothendieck Toposes as Unifying 'Bridges': A Mathematical Morphogenesis

Olivia Caramello

Abstract We present some philosophical principles underlying the theory of topos-theoretic 'bridges', introduced by the author in 2010 and further developed and applied in the subsequent years.

Keywords Unification · 'Bridge' · Invariant · Sheaf · Topos · Site · Equivalence · Duality · Symmetry · Translation

9.1 Introduction

In this paper our aim is to expose some of the philosophical principles underlying our view of Grothendieck toposes as unifying 'bridges' between different mathematical contexts or theories. This view first emerged in Caramello (2010) and was further developed both theoretically (see Caramello 2017) and in relation to specific applications in different fields of Mathematics throughout the past years; see Caramello (2016a) for an overview of the main results obtained so far by applying this methodology. Thanks to these bridges, we can effectively relate – often in profound and unexpected ways – notions, properties and results of different mathematical theories that may well belong to seemingly distant fields and look disconnected at a first glance. In other words, these techniques enable us to multiply,

The author gratefully acknowledges MUR for the support in the form of a "Rita Levi Montalcini" position, Alain Connes and the two anonymous referees for their very useful comments, which have led to an improved presentation of the contents of the paper.

O. Caramello (✉)
Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Como, Italy

Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France
e-mail: olivia.caramello@uninsubria.it; olivia@ihes.fr

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_9

in a sense, points of view on a given problem and to discover hidden relations between distinct mathematical contexts.

A clarification on terminology is needed. In this text, we shall widely use the term 'object' not (only) in the technical sense of category theory, but to refer to any entity, whether abstract or concrete, considered in the context of our methodology. So an 'object' could be a notion, a concept or also a concrete entity belonging to the 'real' world. Similarly, our use of the term 'construction' or 'building' of new objects from given ones should not be intended concretely, as it can refer to an abstract way for associating new objets with given ones (as it is indeed the case in the context of topos-theoretic 'bridges'), not necessarily in a constructive way. When we say 'topos', we always mean 'Grothendieck topos'.

Before proceeding to describe how the 'bridge' technique is implemented in the context of toposes, we shall clarify the sense in which we shall talk about unification and introduce the general concept of a 'bridge' object, as it has been inspired by these mathematical investigations. For examples of 'bridges' of non-mathematical nature, we refer the reader to Caramello (2016b, 2018), while an illustration of selected topos-theoretic 'bridges' is provided by Caramello (2016a).

Lastly, a disclaimer. While the general idea of a 'bridge object', as presented in Sect. 9.3, does not even require a mathematical background to be understood, our comments on its implementation in the context of toposes can hardly be appreciated without a basic knowledge of the language of category theory. As an introduction to this subject, we recommend the classical but still excellent book Mac Lane (1971) by S. Mac Lane. Readers interested in the philosophical and historical aspects of category theory may consult (Krömer, 2007; Marquis, 2010; Mazur, 2008).

9.2 What Does 'Unifying' Mean?

There are two different main significations to the concept of unification. One usually refers to a unifying framework as a general context subsuming a number of particular instances. So, for example, the concept of a category is unifying in the sense that concepts as different as that of a preorder or that of group can all be seen as particular cases of it. Similarly, the language of set theory (in one or another of its standard formalizations) is unifying in that one can ultimately express all mathematical concepts in terms of sets.

Still, one immediately realizes that this kind of unification, based on *generalization*, is *static* in the sense that being able to fit different objects in one context does not provide by itself a means for transferring knowledge between them. For example, knowing that the notions of preorder and of group are both particular cases of the notion of category does not give by itself a tool for transferring results about preorders to results about groups (or conversely), since such results are not necessarily specializations of general results about categories in these two contexts; in fact, the most interesting theorems about preorders or groups are *not* of this form, since they exploit in a non-trivial way the specific characteristics of the given

objects, i.e. precisely the aspects of them that are lost in the generalization! This illustrates the fact that, in a generalization, the diversity of the single objects is not at all valorized, but rather 'forgotten'. In fact, any abstraction process consists in focusing the attention on a limited number of aspects of the given situation at a time, and drawing inferences exclusively on the basis of them, temporarily forgetting all the other ones (see also Sect. 9.11 below for more on abstraction in mathematics).

It is therefore natural to wonder if it is possible to achieve a more substantial form of unification which takes into account the diversity of the single objects in a significant way, rather than 'diluting' it in a generalization, and which is *dynamic* in the sense of allowing non-trivial transfers of information between the objects. One would want a unification that not only valorizes the diversity of the given objects but possibly *explains* its origin in relation to a unity lying somewhere else, as in a sort of *morphogenesis*.

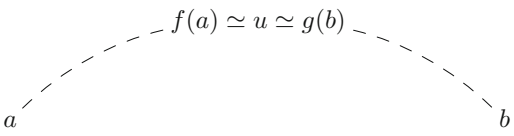
Our (positive) answer to these questions involves the notion of a *bridge object*. We shall discuss for simplicity the construction of 'bridges' across two given objects, even though the technique, as it will be clear, is applicable to arbitrarily big collections of objects.

9.3 The Idea of 'Bridge'

When we compare two objects with each other, we are interested in identifying the *invariants*, that is, the aspects that they have in common despite their diversity. These aspects can sometimes be identified in a concrete way, that is without the need to go out of a (relatively restricted) environmnent in which the objects lie. But most often, it is necessary to adopt novel points of view on the two objects capable of enlightening their (more or less hidden) invariants. This is what 'bridge' objects can achieve.

We can define a *bridge object* connecting two objects a and b as an object u which can be constructed from (or associated with – not necessarily in a concrete sense), from any of the two objects (independently from each other) and admits two different 'presentations' $f(a)$ and $g(b)$ (typically expressing the ways in which u is 'built' respectively from a and from b), in the sense that there is some kind of identification (in mathematics this is typically formalized as an equivalence relation) $\simeq$ both between u and $f(a)$ and between u and $g(b)$.

Such a 'bridge' object u allows us to build 'bridges' allowing *transfers of information* between a and b, as follows. For any property or notion applicable to u which is invariant with respect to the relation $\simeq$, one should attempt to 'translate' it into a property or notion applicable to the object a (resp. to the object b) by using the presentation $f(a)$ (resp. $g(b)$) of the bridge object u in terms of the object a (resp. b).



For instance, in the case of an invariant property P applicable to u, one looks for logical relationships of the kind '$f(a)$ satisfies P if and only if a satisfies a certain property P_a' and '$g(b)$ satisfies P if and only if b satisfies a certain property Q_b'; in such a situation, one gets a logical equivalence between P_a and Q_b, since both properties correspond to the same invariant property P of the bridge object, but understood from the points of view of the objects a and b via the presentations $f(a)$ and $g(b)$ of u. Notice that, since a and b can be objects of very different nature, the properties P_a and Q_b can be concretely very different, in spite of the fact that they represent different manifestations of a unique property, namely P, of the bridge object u.

Our notation $f(a)$ and $g(b)$ for the different presentations of the bridge object u is motivated by the mathematical applications, where these presentations of u are typically the result of applying some functions f and g respectively to a and to b. In general, though, $f(a)$ (resp. $g(b)$) may be any object associated with a (resp. b) by means of some kind of 'assignment', 'procedure' or 'construction' f (resp. g). In most cases, the passage from a (resp. b) to $f(a)$ (resp. $g(b)$) will involve a loss of information about a and b; that is, in general, one will not be able to 'reconstruct' (the whole of) a (resp. b) from $f(a)$ (resp. $g(b)$). Still, some essential information about a and b must be retained in this passage for non-trivial bridges to be established.

Bridge objects are in a sense universal invariants, since, almost tautologically, all the invariants considered on them 'factor' through them. In general, every bridge object supports infinitely many invariants (as any 'genuine' property of a bridge object is, essentially by definition, an invariant). Any invariant allows to transfer different information, behaving like a 'pair of glasses' capable of 'discerning' (or 'unveiling') hidden connections 'coded' in the equivalence between the two presentations of the bridge object. Of course, each of these transfers of information is partial, since each invariant embodies only *some* of the aspects that the two objects have in common. Ideally, in order to achieve a full transfer of information, one would need to consider all the possible invariants, and therefore all the possible bridge objects (and higher-order architectures involving them) connecting the two given objects.

The complexity of the 'unravelings' of properties of the bridge objects in terms of properties of its presentations (when they are at all feasible, in a non-tautological way) may vary enormously from case to case, depending on the sophistication of the invariant considered on u and on that of the constructions of the bridge object u from the two objects a and b. Still, the kind of unification realized by such a method is much more substantial than that achieved by a generalization. Indeed, the diversity of the two objects a and b is no longer 'forgotten', but becomes directly

responsible for the different *forms* in which the invariants defined at the level of the bridge object manifest. So, we have a sort of *morphogenesis* which *explains* the origin of the diversity of different expressions of the same invariant.

Notice that in the 'bridge' technique, it is not the objects themselves to be unified, but their properties, or notions involving them. In other words, the unification takes place at an higher, more abstract level than that of the two objects. In fact, we have used the metaphor of the 'bridge' to underline the fact that we have two distinct levels, that of the objects to be investigated in relation to one another, and that of the bridge objects susceptible of connecting them.

It is important to bear in mind that a 'bridge' object may have a completely different nature from those of the objects which it relates to each other; the two levels do not in general collapse to a single one. The 'flat' bridges (i.e. those in which two levels can be identified) are the relatively uninteresting ones, since they correspond to the situations where the two objects can be directly connected to each other, and in which the unification boils down to 'homogeneisation'. In fact, in the context of theories with the same semantic content connected to each other by their common classifying topos (acting as a 'bridge' object between them), the pairs of theories that can be connected directly to each other are those which are bi-interpretable (in the sense of having equivalent syntactic categories), that is between which there exist 'dictionaries' (provided by the bi-interpretation) allowing to directly relate their syntaxes (cf. section 2.2 of Caramello 2017).

Any 'bridge' connects two levels, which can be thought of as the level of the 'contingent' (or 'concrete'), to which the objects to be unified belong to, and that of the universal (or 'abstract'), where bridge objects and the invariants defined on them lie.

The 'bridge' technique notably provides an approach to the problem of classifying invariants with respect to a given equivalence relation, and obtaining canonical representatives for the equivalences classes. Indeed, suppose that one wants to compare two objects a and b belonging to a set I on which an equivalence relation $\sim_I$ is defined. In such a situation, it is important to identify (and, possibly, classify) the properties of the objects of I that are *invariant* with respect to the relation $\sim_I$, since any such property will allow a transfer of information between a and b. Depending on the cases, this can be an approachable task or an hopelessly difficult one. In fact, a relationship between two given objects is in general an *abstract* entity, which lives in an ideal context which is generally different from the restricted, 'concrete' environment in which one typically considers the two objects (note that, even when the objects themselves are abstract, a relationship between them lies at a higher level of abstraction since it requires a higher degree of logical complexity to be formalized). It thus becomes of crucial importance to identify more 'concrete' (that is, 'easier to represent' or to investigate') entities which could act as 'bridges' connecting the two given objects, by representing in particular their common equivalence class in a form which is as 'concrete' (again, in the sense of ' manageable', or 'easily representable by the human mind') as possible. Think for instance of the complex plane, which is formally defined as the quotient $\mathbb{R}[x]/(x^2 + 1)$, but whose elements can be concretely represented as pairs of real

numbers, or to the set $\mathbb{Q}$ of rational numbers, defined as a quotient of the product $\mathbb{Z} \times \mathbb{Z}^*$ (where $\mathbb{Z}^*$ denotes the set of non-zero integers), whose elements, which are equivalence classes, admit as canonical representatives the reduced fractions whose numerator (or denominator) has a fixed sign.

For this, the concept of *invariant construction* is relevant. We define an invariant construction $f : (I, \sim_I) \to (O, \sim_O)$ between sets I and O on which are defined equivalence relations $\sim_I$ and $\sim_O$, as a function $f : I \to O$ which respects the equivalence relations (i.e., such that whenever $x \sim_I y$, $f(x) \sim_O f(y)$). An invariant construction f is said to be *conservative* if it reflects the equivalence relations (i.e., whenever $f(x) \sim_O f(y), x \sim_I y$). In such a situation, a bridge object connecting two objects $x, y \in I$ will be an object $u \in O$ such that $u \sim_O f(x)$ and $u \sim_O f(y)$. If f is a conservative invariant construction $(I, \sim_I) \to (O, \sim_O)$ then bridge objects in O notably represent equivalence classes of objects of I modulo the equivalence $\sim_I$.

Of course, the usefulness of such bridges greatly depends on whether it is more manageable to work with objects of type O than with objects of type I, or when the relation $\sim_O$ is more tractable than the relation $\sim_I$. Still, experience (both in mathematics and in other sciences) shows that, in the majority of situations, one needs, in order to effectively connect two objects of I, to move from the level of I to the level O of another object being able to serve as a 'bridge' across them (see also Sect. 9.11 for a discussion of this point in the context of the topos-theoretic 'bridge technique).

We anticipate that, in the context of our topos-theoretic 'bridges', the objects a and b to be investigated in relationship with each other will be specific mathematical contexts (represented as sites, theories or other objects from which toposes can be constructed), and $f(a)$ and $g(b)$ will be toposes attached to them which capture a 'common essence'. The 'bridge' technique can be notably applied in the context of theories classified by the same topos, for transferring information across them. Recall (Makkai and Reyes 1977) that any mathematical theory of a very general form (technically speaking, a geometric theory) admits a classifying topos, which, by definition, classifies its models in arbitrary toposes and thus embodies its semantic content (see also Marquis (2010) for an excellent conceptual introduction to categorical logic). Two theories are classified by the same topos (i.e. are *Morita-equivalent*) when, broadly speaking, they describe the same structures in their respective (possibly very different) languages. As we shall see in Sect. 9.6, the construction of the classifying topos defines a conservative invariant construction from the collection of geometric theories (endowed with the notion of Morita equivalence) to that of Grothendieck toposes (endowed with the notion of categorical equivalences of toposes), and classifying toposes can effectively act as 'bridge' objects across Morita-equivalent theories.

Examples of 'bridges' outside mathematics are discussed in Caramello (2016b, 2018).

9.4 Sheaves, or the Passage from the Local to the Global

Grothendieck toposes (Artin et al. 1972) are, by definition, categories (equivalent to a category) of sheaves (of sets) on a site. The notion of site arises from an abstraction of the notion of covering of an open set by a family of open subsets in a given topological space, and represents the most general categorical context for defining sheaves. The notion of sheaf on a topological space was introduced by J. Leray: a sheaf on a topological space X is a way of assigning to each open set U of X a set $\mathcal{F}(U)$ and to each inclusion between open sets $V \subseteq U$ a map $\rho_{V,U} : \mathcal{F}(U) \to \mathcal{F}(V)$ in such a way that $\rho_{U,U} = 1_{\mathcal{F}(U)}$ for each U and $\rho_{W,V} \circ \rho_{V,U} = \rho_{W,U}$ for each $W \subseteq V \subseteq U$ (these are the conditions that define the notion of *presheaf* on X) plus the requirement that, for any open covering of U by a family $\{U_i \mid i \in I\}$ of open subsets $U_i \subseteq U$, giving an element of $\mathcal{F}(U)$ corresponds precisely to giving a family $\{x_i \in \mathcal{F}(U_i) \mid i \in I\}$ of elements of the $\mathcal{F}(U_i)$'s which is compatible in the sense that $\rho_{Z,U_i}(x_i) = \rho_{Z,U_j}(x_j)$ for each $i, j \in I$ and any open set Z contained both in U_i and in U_j.

The canonical example of a sheaf on a topological space X is that of continuous real-valued functions, which assigns to each open set of X the set of continuous real-valued functions on U and to each inclusion $V \subseteq U$ between open sets the operation of restriction of such functions on U to V.

Categorically, a presheaf is simply a functor from the opposite of the category $\mathcal{O}(X)$ of open sets of X (whose objects are the open sets of X and whose arrows are the inclusions between them) to the category of sets.

Sheaves on a topological space X can be defined as presheaves satisfying the above gluing condition, which can be expressed categorically entirely in terms of the category $\mathcal{O}(X)$ and of the notion of covering family $\{U_i \hookrightarrow U \mid i \in I\}$ in this category.

A (small) site is a pair $(\mathcal{C}, J)$ consisting of a (small) category $\mathcal{C}$ and a so-called *Grothendieck topology* J on it, which specifies a notion of 'covering family' of arrows in $\mathcal{C}$, with respect to which one can formulate a sheaf condition.

A presheaf on a category $\mathcal{C}$ is simply a contravariant functor from $\mathcal{C}$ to the category **Set** of sets. Given a Grothendieck topology J on $\mathcal{C}$, a presheaf P on $\mathcal{C}$ is said to be a *J-sheaf* if it satisfies the glueing condition with respect to every compatible family of elements of P indexed by a J-covering family; see Mac Lane and Moerdijk (1992) for the details. The category of J-sheaves on $\mathcal{C}$ and natural transformations between them is denoted by $\mathbf{Sh}(\mathcal{C}, J)$. A Grothendieck topos is any category $\mathcal{E}$ equivalent to the category $\mathbf{Sh}(\mathcal{C}, J)$ of sheaves on a small site $(\mathcal{C}, J)$.

The notion of sheaf expresses a very robust and harmonious relationship between the local and the global. It formalizes the process by which one defines a global entity by specifying its local behaviour on objects covering its domain. For instance, one can define a continuous real-valued function on an open set U of a topological space X by pasting together continuous real-valued functions $f_i : U_i \to \mathbb{R}$ defined on sub-open sets U_i covering U which are compatible with each other (in the sense that they agree on the intersections of the U_i's).

Notice that every set of local data which comes from a global entity by a process of 'instantiation' satisfies some coherence conditions; for example, given a continuous function f defined on an open set U of a topological space, if we define f_V as the restriction of f to an open subset $V \subseteq U$ and f_W as the restriction of f to an open subset $W \subseteq U$, then we have the coherence relation $f_V|_Z = f_W|_Z$ for any open subset $Z \subseteq V \cap W$. The definition of sheaf requires the converse also to hold: any set of compatible local data should uniquely determine a globally-defined *datum*.

The theory of descent data is another (higher-categorical) illustration of the same principle of defining global entities by gluing compatible sets of local data.

Although sheaves might appear very abstract at first sight, they are actually real in a very strong sense. What is reality if not a sheaf of coherent perceptions? Note that we tend to call 'real' anything which is 'independent from' (in the sense of 'invariant with respect to') the perceptions that we might have of it. The reason why we believe in the existence of reality (if we do) is that we continuosly experience coherence relations existing between the perceptions of different individuals or measuring instruments; it is therefore scientifically reasonable (as a minimalist explanation) to suppose the existence of something which would 'generate' (by its mere existence) all these perceptions and hence which would be ontologically responsible for the coherence relations between them.

9.5 Grothendieck Toposes

A general sheaf is by itself a rather rich entity since is specifies not only a fixed, global *datum* but a whole collection of local data which are compatible with each other. Considering all sheaves on a given site to form a category, namely a Grothendieck topos, adds even more coherence relations arising from the 'interaction' between different sheaves. It is thus clear that a Grothendieck topos is an extremely rich entity. Indeed, any topos is a veritable mathematical universe within which one can do mathematics and in particular consider models of arbitrary first-order (and even higher-order) theories.

As a mathematical universe, every topos has its own internal logic, which in general is not classical but intuitionistic and with multiple truth values, reflecting the fact that the notion of truth accommodated by sheaves is local and variable (according to the domain of a generalized element), rather than global and fixed. As far as its internal structure is concerned, a Grothendieck topos satisfies all the completeness properties one might hope for: every universal problem (expressible in terms of existence of limits or colimits) has a solution in a topos, because of its categorical completeness and cocompleteness. Moreover, any topos has exponentials (which are the categorical analogue of the sets of functions from one given set to another), a subobject classifier, which encodes much of its internal logic, separating sets and coseparating sets.

These categorical properties have a number of remarkable consequences: every functor between toposes which preserves all limits (resp. colimits) has a left (resp.

right) adjoint; in particular, every covariant (resp. contravariant) functor from a Grothendieck topos to the category of sets which preserves limits (resp. which sends colimits to limits) is representable. All of this shows that toposes are ideal concepts to be used for building unifying 'bridges' across different mathematical contexts since they 'accommodate' (in the sense of being natural homes for) many objects arising as the solution of universal problems; a typical example is the notion of universal model of a geometric theory within its classifying topos. Moreover, this very rich categorical structure is responsible for the fact that toposes are full of symmetries or, in other words, that they naturally support a great number of invariants.

Grothendieck toposes are also stable with respect to many significant operations that one might want to perform on them; the 2-category of toposes is itself very rich. In particular, the theory of Grothendieck toposes supports relativisation techniques, since the topos of sheaves on a site internal (or relative) to a base Grothendieck topos is again a Grothendieck topos. Being able to change the base topos according to one's needs and to switch from an 'internal' to an 'external' point of view when dealing with toposes in relation to one another is a classical technique that adds further power to the theory (much as Grothendieck's relativisation techniques have played a key role in his refoundation of algebraic geometry in the language of schemes) and naturally leads to the discovery of a great number of different presentations for the same topos.

In a sense, the ontology of toposes is very large. Every mathematical theory, even a contradictory one, finds its home in the context of toposes (note that this is not true in the restricted context of sets, where a contradictory theory does not have any models); in fact, a geometric theory is contradictory if and only if its classifying topos is trivial (in the sense that it is the topos **1** having one object and one arrow on it). Note that, whilst being trivial, this topos does indeed contain a model of the theory, namely its universal model. This is very relevant both from a conceptual and a technical viewpoint; indeed, not having to worry whether something exists allows for a much greater technical flexibility. The problem is no longer whether the object we would like to construct exists or not; in the world of toposes, in a sense, all problems (of a specified but very general kind – think for instance of the existence of limits or colimits for small diagrams) admit a solution, so the 'absolute' problem of the existence of a desired entity gets reduced to a 'relational' problem, that of whether we can transport the 'universal', topos-theoretic solution to a set-theoretic or more concrete structure suitable for our needs (think, as an example, of the topos-theoretic construction of forcing models for set theory). Notice that this very large ontology manifests itself both at the level of the individual objects of toposes, namely sheaves and at the global level of the entire universe of sheaves on a given site, that is, at the level of a given topos, in addition to the level of the (very big) 2-category of toposes.

The 'completeness' of the world of toposes is also reflected in the fact, already hinted at above, that all the functors between toposes that satisfy the necessary conditions for being representable (resp. for admitting a left or right adjoint) *are* indeed representable (resp. *do admit* such a left or right adjoint).

Another very relevant aspect of toposes is their amenability to computation. Any Grothendieck topos is, in a sense, a mathematical environment without 'holes', by virtue of the completeness properties it enjoys; this makes it very convenient for calculations, since one can exploit all its internal symmetries for carrying them smoothly and effectively, never having to worry whether the result of this or that categorical operation exists or not. On the other hand, one does not go far astray by computing in a topos, since, following the 'bridge' philosophy, one can always interpret the results of these computations in terms of relevant presentations for the given topos.

9.6 The Yoneda Paradigm

Recall that a functor from a category $\mathcal{C}$ to a category $\mathcal{D}$ is a way of assigning to each object of $\mathcal{C}$ an object of $\mathcal{D}$ and to each arrow of $\mathcal{C}$ an arrow of $\mathcal{D}$ so as to respect identity arrows and the operations of domain, codomain and composition of arrows.

We can think of a functor as a carrier of information which is indexed by the objects and arrows of the domain category. In particular, a presheaf P on a category $\mathcal{C}$, which by definition is a contravariant functor from $\mathcal{C}$ to the category **Set** of sets, sends any object c of $\mathcal{C}$ to a set $P(c)$ and any arrow $f : d \to c$ in $\mathcal{C}$ to a map $P(f) : P(c) \to P(d)$. As such, a presheaf is in general a carrier of a significant amount of information. Any object c_0 of the category $\mathcal{C}$ determines a presheaf on $\mathcal{C}$, denoted by $\mathrm{Hom}_{\mathcal{C}}(-, c_0)$, which sends an object c of $\mathcal{C}$ to the collection $\mathrm{Hom}_{\mathcal{C}}(c, c_0)$ of arrows in $\mathcal{C}$ from c to c_0 and sends an arrow $f : d \to c$ in $\mathcal{C}$ to the function $\mathrm{Hom}_{\mathcal{C}}(c, c_0) \to \mathrm{Hom}_{\mathcal{C}}(d, c_0)$ between these home sets given by composition with f on the right.

Recall that a presheaf P is said to be *representable* if it is (up to isomorphism) of the form $\mathrm{Hom}_{\mathcal{C}}(-, c_0)$ for some object c_0 of $\mathcal{C}$. This means that there is an element $x_0 \in P(c_0)$ which 'generates' all the elements of P in the sense that for any object c of $\mathcal{C}$ and any element $x \in P(c)$ there is a unique arrow $c \to c_0$ such that $P(f) : P(c_0) \to P(c)$ sends x_0 to x. When a functor is representable, this means that all the information carried out by it is actually concentrated in a single object, namely the pair (c_0, x_0) representing it. If the functor carries a lot of information, proving its representability is a significant result, since it shows that the functor has a sort of 'center of symmetry', given by its representing pair, which generates by 'deformation' all its elements associated with arbitrary objects of the category.

The assignment $c_0 \mapsto \mathrm{Hom}_{\mathcal{C}}(-, c_0)$ can actually be made into a functor

$$y_{\mathcal{C}} : \mathcal{C} \to [\mathcal{C}^{\mathrm{op}}, \mathbf{Set}],$$

called the *Yoneda embedding* of $\mathcal{C}$ into the category $[\mathcal{C}^{\mathrm{op}}, \mathbf{Set}]$ of presheaves on $\mathcal{C}$. This functor is full and faithful; in particular, it reflects isomorphisms. This shows that an object c can be identified with the corresponding representable functor $\mathrm{Hom}_{\mathcal{C}}(-, c)$, whose elements are the arrows from arbitrary objects of $\mathcal{C}$ to c. These

arrows are called the *generalized elements* of c; this terminology is justified by the fact that, in the category **Set** of sets, any element of a set c can be identified with an arrow $1 \to c$ in **Set**, where 1 is the singleton set $\{*\}$. A generalized element of an object c, as an arrow to c, thus defines a point of view on c, or better a 'direction of observation' of c within the category $\mathcal{C}$. We can thus interpret the above result by saying that an object can be identified with its generalized elements, that is, broadly speaking, with the collection of points of view that we can have on it. We call this the *Yoneda paradigm*.

Notice that, in this context as well, there are coherence relations between the points of view that one can have on a given object: for instance, any arrow $b \to a$ in $\mathcal{C}$ canonically induces a way of mapping the generalized elements of c defined on b to the generalized elements of c defined on a. These notions suggest that, in general, whenever one experiences coherence relations, one should look for the 'source' of them, possibly in the form of a representing pair for a functor encoding such relations (see also Sect. 9.7 below). As we argued in Sect. 9.4, the language of sheaves is particularly apt to formalize local-global coherence relations. Grothendieck had the key idea of considering *all* sheaves on a given site to form his toposes, supported by the conviction that, since each of them behaves as a sort of " meter[1]" of the site, it is all the more powerful to consider a measure instrument not in an isolated way but in connection with all the other measure instruments that one might want to dispose of. This in fact leads to a whole universe of coherence relations, namely a topos.

The identification of the objects c of a category $\mathcal{C}$ with their functors $\mathrm{Hom}_{\mathcal{C}}(-, c)$ of generalized elements realized by the Yoneda embedding is very relevant also from a technical viewpoint, since it allows to understand constructions internal to the category $\mathcal{C}$ in set-theoretic terms. For example, since the Yoneda embedding $y_{\mathcal{C}} : \mathcal{C} \to [\mathcal{C}^{\mathrm{op}}, \mathbf{Set}]$ preserves and reflects all limits, limits in $\mathcal{C}$ can be understood in terms of limits in the corresponding presheaf topos $[\mathcal{C}^{\mathrm{op}}, \mathbf{Set}]$, which are in turn calculated componentwise in terms of the relevant limits in **Set**.

The Yoneda lemma has a beautiful incarnation in the context of toposes, which provides a further illustration of their natural symmetries and completeness properties. Every Grothendieck topos $\mathcal{E}$ can be endowed with a Grothendieck topology $J_{\mathcal{E}}^{\mathrm{can}}$, called the canonical one, whose covering sieves are those which contain small epimorphic families; this is the largest Grothendieck topology on $\mathcal{E}$ for which all the representable functors are sheaves. Now, the Yoneda embedding $y_{\mathcal{E}} : \mathcal{E} \to [\mathcal{E}^{\mathrm{op}}, \mathbf{Set}]$ yields an equivalence

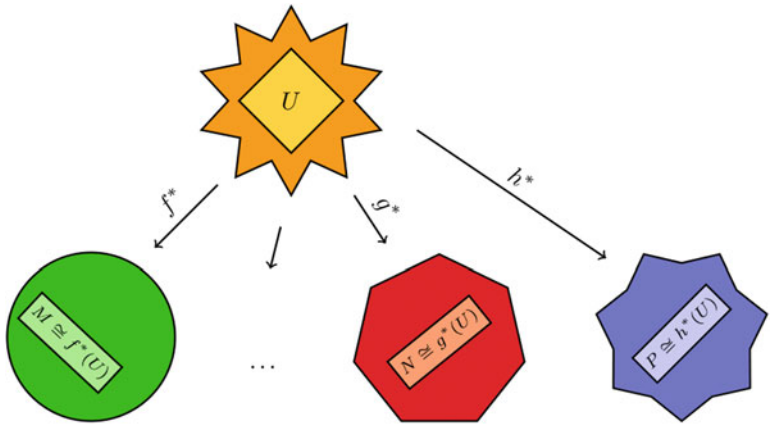
$$\mathcal{E} \simeq \mathbf{Sh}(\mathcal{E}, J_{\mathcal{E}}^{\mathrm{can}})$$

between $\mathcal{E}$ and the category of sheaves on this (large in general, but always small-generated) site. This result can be profitably applied not only for understanding

[1] See section 2.12 of his work *Récoltes et Semailles*.

limits in $\mathcal{E}$, but also for describing non-trivial constructions in $\mathcal{E}$, such as colimits, in terms of generalized elements of objects of $\mathcal{E}$.

Finally, we remark that the theme of representability plays a key role in connection with the topos-theoretic 'bridge' technique. Indeed, the classifying topos of a geometric theory can be defined as a representing object for the (pseudo)functor of models of the theory; its generalized elements (within the category of Grothendieck toposes) are the categories of models of the theory in arbitrary toposes. The Yoneda paradigm thus tells us that a Grothendieck topos can be identified with the (pseudofunctor of) structures that it classifies. In particular, it contains a distinguished model of the theory, called its *universal model*, which 'generates' all the models of the theory in arbitrary toposes:



Classifying topos

In the picture, the big coloured shapes represent different toposes while the inner lighter shapes represent models of a given theory inside them; in particular, the dark yellow star represents the classifying topos of the theory and the light diamond represents 'the' universal model of the theory inside it. The classifying topos thus resembles to a 'sun' generating shadows in all directions; when we look at particular models of the theory in toposes, we are just contemplating deformations of this universal model by means of structure-preserving functors (technically speaking, inverse images of geometric morphisms of toposes), a bit as if we were in Plato's cave.

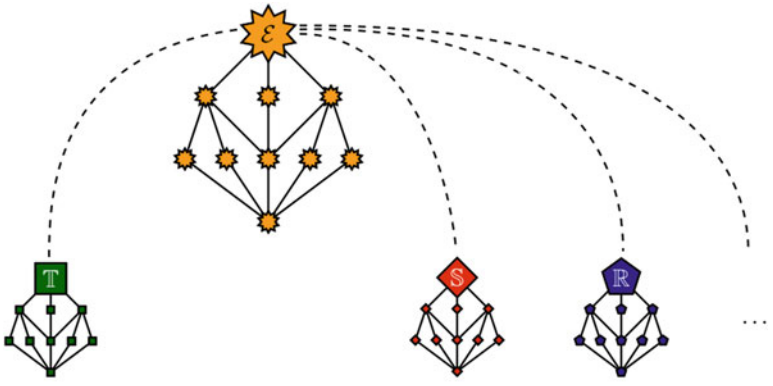
By definition of classifying topos, two theories are Morita-equivalent (i.e. they have equivalent classifying toposes) if and only if they have equivalent categories of models in any topos, naturally in the topos, that is, broadly speaking, if they describe, each in its own language, the same structures, or, in other words, if they have the same mathematical content (embodied by the common classifying

topos). Notice that the classifying topos construction constitutes a conservative invariant construction (in the sense of Sect. 9.3) from the world of geometric theories endowed with the relation of Morita equivalence to the world of Grothendieck toposes endowed with the relation of categorical equivalence; one can then understand, given the fact that the notion of categorical equivalence is much more technically tractable than Morita equivalence (for instance, one can easily see whether a property is a categorical invariant, and introduce infinitely many new categorical invariants without any effort), the crucial role that classifying toposes can play as 'bridges' across Morita-equivalent theories.

9.7 Generation from a Source

There are different senses in which one can understand the idea of 'generation from a source' which we hinted at in Sect. 9.4. We shall be concerned in particular with two of them. The first is the Yoneda paradigm of representable functors, discussed in Sect. 9.6; the second, which is more subtle as it lies at a higher level of abstraction, is the relationship between a bridge object and its different 'representations'. This duality is akin to that between a theme and a number of variations on it: anything which happens at the level of the bridge object will have 'ramifications' in the context of all its representations. Such 'ramifications' will then entertain coherence relations between them just as different 'variations' on the same 'theme' lying at the level of the bridge object, which then appears as ontologically 'responsible' for the existence of such relations.

For instance, the following picture represents the lattice structure on the collection of the subtoposes of a topos $\mathcal{E}$ inducing lattice structures on the collection of 'quotients' (i.e. geometric theory extensions over the same signature) of geometric theories $\mathbb{T}$, $\mathbb{S}$, $\mathbb{R}$ classified by it.



Lattices of theories

Notice that there are infinitely many theories classified by the same topos, which may belong to different areas of mathematics. So, if one ignores toposes, it would be difficult to realize that a particular structure we are dealing with in a specific mathematical context has in fact a counterpart in another field of mathematics just because this structure is induced by a universal structure lying at the topos-theoretic level. It would therefore be of great usefulness for the 'working mathematician' to realize about which of the construction (s)he deals with can be lifted at the topos-theoretic level, so that alternative versions of them can be obtained by switching to different representations of the same topos. Indeed, a topos is an object that, by embodying, in a sense, the collection of all points of view on a given topic, represents a crossroads between different mathematical paths, a place where different perspectives and languages converge mirroring one into another.

By making a topos-theoretic analysis of the concepts one works with, one is also able to understand whether the notions one is dealing with are sufficiently *robust* or *modular* (in the sense that they correspond to topos-theoretic invariants), in which case they admit infinitely many reformulations in different contexts providing multiple points of view on the given topic, or whether instead they are ad hoc, *concrete* notions that perhaps serve a very specific purpose but which occupy a relatively marginal place in the mathematical landscape.

In our context, there are also two main meanings that we can give to the expression 'point of view'. As argued in Sect. 9.6, we can think of a generalized element of an object in a category as a point of view on it. But it is also natural to think of a presentation of a bridge object as a point of view on it. In fact, in the context of classifying toposes, their different presentations correspond to different theories classified by them, which indeed provide, each with its own language, different points of view on these toposes.

When we talk about *morphogenesis* (see Sect. 9.3), we refer to the fact that different invariants on a bridge object may manifest themselves in different ways in the context of different presentations of that object. We can interpret this by saying that every form that exists abstractly at the level of a bridge object differentiates giving rise to different forms in the context of different presentations of that object. It is this process of 'differentiation from the unity' that we call 'morphogenesis'. As an example of the significantly different ways in which even basic invariants of toposes manifest in the context of different sites, consider the property of a topos to be bivalent: in the context of a trivial site $(\mathcal{C}, T)$, it corresponds to the property of $\mathcal{C}$ to be strongly connected (in the sense that for any objects a and b of $\mathcal{C}$, there is both an arrow $a \to b$ and an arrow $b \to a$), in the context of an atomic site $(\mathcal{C}, J_{\mathrm{at}})$ it corresponds to the property of $\mathcal{C}$ to be non-empty and to satisfy the joint embedding property (that is, the property that any two objects of $\mathcal{C}$ admit an arrow from a third one), and in the context of the classifying topos of a geometric theory $\mathbb{T}$ (represented as the category of sheaves on its syntactic site) it corresponds to the property that $\mathbb{T}$ be geometrically complete (in the sense that every geometric assertion in the language of the theory is either provably true or provably false in it, but not both). As another example, the property of a topos to satisfy De Morgan's law manifests in the context of a topos of the form $[\mathcal{C}, \mathbf{Set}]$ as the amalgamation

property on $\mathcal{C}$ (i.e. the property that every pair of arrows with common domain can be completed to a commutative square), while in the context of the topos of sheaves $\mathbf{Sh}(X)$ on a topological space it corresponds to the property of X to be extremally disconnected (in the sense that the closure of any open set is open). As yet another example, take the property of a topos to be Boolean; in the context of a presheaf topos $[\mathcal{C}^{\mathrm{op}}, \mathbf{Set}]$ it corresponds to the property of $\mathcal{C}$ to be a groupoid, while in the context of the topos of sheaves $\mathbf{Sh}(X)$ on a topological space it corresponds to the property of X to be almost discrete. Notice that these are relatively simple invariants of toposes that are very easily calculated; still, their different manifestations in the context of different presentations are very surprising (without the point of view of toposes, it would have been hard to imagine that they could be related to each other), which gives an idea of the mathematical morphogenesis induced by topos-theoretic invariants (which of course will be much greater than in the above examples in the case of more sophisticated invariants).

It is important to remark that, whilst 'concretely' very different, all the manifestations of a given invariant in the context of different presentations are *compatible* with each other and actually entertain coherence relations ultimately resulting from the fact that they all come from the same source. Most strikingly, this morphogenesis is entirely determined by the structural relationship between a topos (or, more generally, a bridge object) and its different presentations.

As an example of a 'bridge' obtained by using the above-mentioned invariants, we mention our topos-theoretic interpretation (Caramello 2014) of Fraïssé's theorem in Model Theory, where the classifying toposes of the theories $\mathbb{T}'$ of homogeneous $\mathbb{T}$-models (where $\mathbb{T}$ is a geometric theory classified by a presheaf topos whose category f.p.$\mathbb{T}$-mod($\mathbf{Set}$) of its finitely presentable models satisfies the amalgamation property) are presented, on the one hand, as the categories $\mathbf{Sh}(\text{f.p.}\mathbb{T}\text{-mod}(\mathbf{Set})^{\mathrm{op}}, J_{\mathrm{at}})$ of sheaves on f.p.$\mathbb{T}$-mod($\mathbf{Set}$)$^{\mathrm{op}}$ with respect to the atomic topology and, on the other hand, as the categories $\mathbf{Sh}(\mathcal{C}_{\mathbb{T}'}, J_{\mathbb{T}'})$ on the geometric syntactic site of $\mathbb{T}'$. Transferring the invariant property of being bivalent across these two presentations yields the following result: $\mathbb{T}'$ is complete if and only if f.p.$\mathbb{T}$-mod($\mathbf{Set}$) is non-empty and satisfies the joint embedding property. Notice that completeness is in general a hard-to-establish property of theories, while the joint embedding property, at least in a great number of situations, is a much more 'concrete' and easier property to investigate.

9.8 Sites and Toposes, or the Contingent and the Universal

The key element on which the 'bridge' technique is based is the fundamental ambiguity consisting in the fact that any given topos has infinitely many different presentations. Topos-theoretic invariants can thus be used for generating 'bridges' connecting such presentations. As in any 'bridge', we have two levels: that of the contingent, represented in this case by sites, axiomatic presentations of theories or

other objects by means of which toposes can be presented, and that of toposes, which is the abstract level where invariants naturally live.

Every topos-theoretic invariant generates a veritable mathematical morphogenesis resulting from its expression in terms of different presentations of toposes, which often gives rise to connections between 'concrete' properties or notions that are completely different and apparently unrelated from each other.

The mathematical exploration is therefore in a sense 'reversed' with respect to the more classical, 'bottom-up' approaches since it is guided by the equivalences between different presentations of the same topos and by topos-theoretic invariants, from which one proceeds to extract concrete information on the theories or contexts that one wishes to study.

Toposes can be thought of as 'stars' that enlighten mathematical reality (cf. Sect. 9.6), as universal standpoints on theories which naturally unveil their symmetries. A site, or, more generally, an object from which a topos can be built, is, in a sense, a point of view on that topos. Any site $(\mathcal{C}, J)$ can be 'mapped' to the corresponding topos by means of the canonical functor

$$l : \mathcal{C} \to \mathbf{Sh}(\mathcal{C}, J)$$

to the topos of sheaves on it. Interestingly, there are infinitely many 'intermediate' sites between it and the associated topos (which can be considered itself a site, by equipping it with the canonical topology C), obtained by equipping any full subcategory of $\mathbf{Sh}(\mathcal{C}, J)$ containing the image of $\mathcal{C}$ with the Grothendieck topology induced by C. These sites can be viewed as different 'scales' of observation for phenomena formalized as topos-theoretic invariants; the way in which such invariants express in terms of them would then account for the existence of multiple descriptions of invariant laws at different scales. Having different and apparently incompatible descriptions for a given 'physical' content at different scales is a fundamental problem that physicists face; a topos-theoretic analysis of this kind of problems could therefore be highly beneficial.

A simple example illustrating this last remark is provided by the construction of the Alexandrov space $\mathcal{A}_{\mathcal{P}}$ associated with a preorder $\mathcal{P}$. Recall that $\mathcal{A}_{\mathcal{P}}$ is the topological space whose underlying set is $\mathcal{P}$ and whose open sets are the upper sets with respect to the preorder relation, that is the subsets U of $\mathcal{P}$ such that whenever $a \leq b$ in $\mathcal{P}$, $a \in U$ implies $b \in U$. Now, it is easy to see that we have a canonical equivalence

$$[\mathcal{P}, \mathbf{Set}] \simeq \mathbf{Sh}(\mathcal{A}_P) \ .$$

The first presentation $[\mathcal{P}, \mathbf{Set}]$ of this topos is in a sense 'combinatorial', while the second is topological; indeed, the objects of the first site are the elements of $\mathcal{P}$, considered as objects of a category and hence deprived of internal structure (as if they were elementary particles), while the objects of the canonical site presenting $\mathbf{Sh}(\mathcal{A}_P)$ are open sets of $\mathcal{P}$, which instead have a rich internal structure supporting a geometrical intuition. Accordingly, when a given topos-theoretic invariant is studied

from the point of view of these two presentations, in the first case one obtains a combinatorial formulation of it, while in the second a topological formulation. Take for instance the property of a topos to be De Morgan; as we saw in Sect. 9.7, this reformulates in terms of $\mathcal{P}$ as the amalgamation property on it (that is, the property that for any elements $a, b, c \in \mathcal{P}$ such that $a \leq b$ and $a \leq c$ there is $d \in \mathcal{P}$ such that $b \leq d$ and $c \leq d$) and in terms of $\mathcal{A}_{\mathcal{P}}$ as the property of this space to be extremally disconnected.

Let us further elaborate more generally on this duality between the contingent and the universal.

Every language or point of view is partial and incomplete (i.e. full of 'holes'), and it is only through the integration of all points of view that we can capture the essence of things (cf. Sects. 9.7 and 9.6).

There is no universal language that would be better (in an absolute sense) than all the others; every point of view highlights certain aspects and hides others and can be more convenient than another in relation to a certain goal. Universality should thus be researched not at the level of languages (or 'points of view') but at the level of the 'ideal' objects on which invariants are defined.

It is therefore necessary to reason at two levels, that of the invariants (and the 'bridge' objects on which they are defined) and that of their manifestations in the context of 'concrete' situations, and to study the duality between these two levels, a duality that can be thought of as the one between a 'sense' and the different ways to express it. These two levels are independent from each other and important each in its own right; as observed in Sect. 9.3, confusing them makes unification collapse to 'homogenisation'.

9.9 Invariant-Based Translations

The 'bridge' technique is a methodology for translating notions, ideas and results across different mathematical contexts. It is important to realize that, in general, such translations are *not* literal, since they are determined by the expression of topos-theoretic invariants in terms of different presentations of toposes, rather than by the use of a 'dictionary'. In fact, as already remarked in Sect. 9.3, 'dictionaries', or direct ways of relating two given objects with each other, exist only in a minority of cases, which in fact are relatively uninteresting since the resulting translations do not essentially change the syntactical shape in which the information is presented and hence do not really generate novel points of view on it; these are precisely the situations where unification collapses to homogenisation.

In fact, even in Linguistics, a good translation is most often a non-literal one; such a translation should be based on a preliminary identification of the invariants, that is of the aspects of the text which one wants to remain unchanged (i.e. preserved) in the transition process from one language to the other. In the case of translations between natural languages, the most obvious invariant is meaning, but there are others too: for instance, one might also, or in alternative, want to preserve a particular type

of metre or musicality, especially when translating a poem. Anyway, whatever the invariants, what matters is to let them guide the translation, that is play the role of bridge objects determining the translation by virtue of how they express in the two different languages. The same happens with the mathematical translations based on topos-theoretic 'bridges': in this case the objects to be related are mathematical contexts or theories and the bridge objects are toposes associated to them, on which an infinite number of invariants are defined.

9.10 Symmetries by Completion

As a matter of fact, the more one enlarges the mathematical environment where one works, the higher is the number of internal symmetries that one generates. Think for example of number systems; the development of mathematics has gone progressively in the direction of extending them in order to find solutions to certain problems (such as finding inverses to certain naturally defined operations such as addition, multiplication, taking powers etc.). For instance, from the set of natural numbers one has constructed the integers by formally adding negative numbers: by doing this, a fundamental symmetry with respect to the zero has appeared. Similarly, by passing from the real line to the complex plane in order to notably find a solution to the equation $x^2 + 1 = 0$, one has found a much more 'symmetric' mathematical environment, as witnessed in particular by the fundamental theorem of algebra (which provides a perfect symmetry between the degree of a polynomial in one variable and the number of its roots counted with their multiplicities, and which has no natural analogue in the restricted context of the real line).

On the other hand, to relate different languages or points of view to each other it is necessary to *complete* them to objects that realize *explicitly* the *implicit* hidden in each of them and which therefore can act as bridge objects connecting them. Indeed, it is at the level of these completed objects that invariants, that is, symmetries, manifest, and that we can understand the relationships between our given objects thanks to the 'bridges' induced by the invariants.

We can see all these principles incarnated by the use of toposes as 'bridges'. As the complex plane $\mathbb{C}$ is obtained from the real line by means of a formal construction (namely, $\mathbb{R}[x]/(x^2 + 1)$) consisting in the addition of certain 'imaginaries', so the topos $\mathbf{Sh}(\mathcal{C}, J)$ of sheaves on a site $(\mathcal{C}, J)$ represents a completion of $\mathcal{C}$ by the addition of imaginaries (indeed, any object of $\mathbf{Sh}(\mathcal{C}, J)$ can be canonically expressed as a 'definable' quotient of a coproduct of objects coming from $\mathcal{C}$). Also, the classifying topos of a theory is constructed by means of a completion process (of the theory itself), with respect, in a sense, to all the concepts that she is potentially capable to express. Thanks to the 'bridge' technique, different theories that describe the same mathematical content are put in relation with each other as if they were fragments of a single object, partial languages that complement each other by mirroring each into one another within the totality of points of view embodied by the classifying topos.

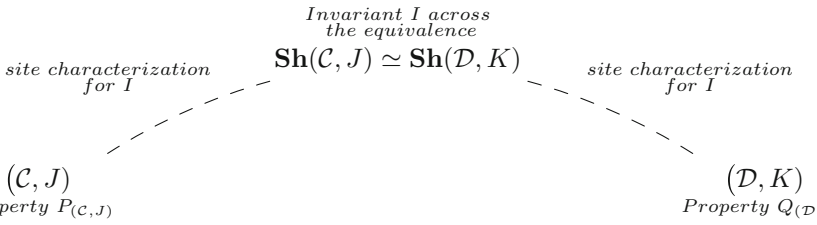
In fact, a translation should not be thought of as a way for relating entities that are necessarily very different from each other, but as a process of discovery of new potential implicit in a certain point of view or language. Every language, in an attempt to express a reality that is much richer, can be compared to a sketch drawn by an artist. In looking at it, our brain operates a sort of automatic completion of it which allows us to understand its implicit meaning. The transition from a linguistic expression to its meaning can thus be thought of as a kind of completion.

9.11 The 'Bridge' Technique in Topos Theory

As anticipated in the previous sections, the construction of a topos-theoretic 'bridge' involves first of all the identification of an equivalence between two different presentations for the same topos; this will form the 'deck' of the bridge, while the different objects presenting the toposes will constitute the extremes of the bridge (to be related with each other). Of course, one can also consider, as 'decks' of bridges, more general kinds of relationships between toposes that are not equivalences, but, whilst invariants under equivalences of categories are easily generated and identified, there are many less invariants available in the case of a relation that is not an equivalence. So, even if one starts with two toposes between which there is a relation that is not an equivalence, it is often convenient to try to modify the toposes involved by suitable operations in order to obtain an equivalence of toposes and then apply the 'bridge' technique to the latter.

Next, one considers a topos-theoretic invariant, and tries to 'unravel' it both in terms of the first presentation and in terms of the second (in a non-tautological way, that is obtaining genuine, 'concrete' expressions for it in the 'languages' of the two presentations); provided that this is feasible, these characterizations will give the two 'arches' of our bridge.

For example, a 'bridge' between different sites of presentations for the same topos induced by an invariant property of toposes, has the following form:



Here the properties $P_{(\mathcal{C},J)}$ and $Q_{(\mathcal{D},K)}$ of the sites $(\mathcal{C}, J)$ and $(\mathcal{D}, K)$, shown by the 'bridge' to be equivalent, are *unified* as different manifestations, in the context of the sites $(\mathcal{C}, J)$ and $(\mathcal{D}, K)$, of the same invariant I lying at the topos-theoretic level.

In practice, the choice of the invariant(s) will depend on the kind of information that one is interested to extract from the equivalence of toposes; indeed, the invariant should be chosen so that its expressions in terms of the different presentations directly relates to the aspects of the problem that one is interested to investigate. Normally, in order to extract a significant amount of information on a non-trivial mathematical problem, one single invariant will not suffice; one will have to consider several invariants (and hence generate several 'bridges') and combine the insights obtained thereby in order to eventually arrive at a global and deep understanding of the problem. In fact, each invariant will allow to 'read' (or 'decode') certain information 'hidden' (or 'coded') in the equivalence of toposes; there is not in general a privileged invariant that would subsume all the others, in the sense that the result generated by it through the 'bridge' technique (in the context of the given equivalence) would entail the results generated by the other invariants.

Any 'bridge' results in a connection between 'concrete' properties or notions involving the objects used for presenting the topos. Indeed, in spite of their crucial role in establishing such a correspondence, toposes do not appear in the final formulation of the result. It is important to bear in mind that the majority of correspondences, dualities and equivalences existing in mathematics are actually 'hidden' (in the sense that they are not induced by 'dictionaries' such as functors between sites or interpretations between theories and hence cannot be fully appreciated 'concretely') and manifest themselves only at the topos-theoretic level; our topos-theoretic reinterpretation and generalization of Galois theory (Caramello 2016c), reviewed in Caramello (2016a), is a compelling illustration of this (see also Sect. 9.12 below).

We should pause to note that this methodology represents a distinctly abstract way of doing mathematics, in the sense that it is an implementation of the principle according to which in order to obtain specific, 'concrete' information about a given mathematical problem, one should abstract it in several different directions (where by abstracting we mean focusing on a limited number of aspects at a time, temporarily forgetting about all the other ones) and then proceed to combine and integrate the insights obtained by investigating the resulting generalisations (or collecting information about them) to derive 'specific', 'concrete' results on the original problem. The underlying idea is that concreteness can be obtained in a top-down way by intersecting abstract planes, much as we can obtain a point by intersecting two lines in a plane or three planes in the three-dimensional affine space. The advantages of such an abstract approach are multiple. First of all, such a process creates a whole web of relations surronding the given problem, generating a sort of 'rain of results' falling into that territory. Moreover, it enlightens the general architecture of the proof and where and how the hypotheses come into play. This leads to a form of *modularity*: since the role of the hypotheses is enlightened, one is in the position to understand how different hypotheses could lead to different results, thereby also realizing a form of *continuity*. In other words, by setting the given problem within a family of related problems, this methodology allows to see it not in an isolated way but as part of a bigger picture, which greatly enhances one's understanding. Still, the 'bridge' technique is radically different from the

traditional, relational way of doing mathematics based on category theory. The difference between our approach and the classical categorical one is particularly apparent in the treatment of duality theory (see Caramello (2020) for a discussion on this). Indeed, the fundamental relational notion in category theory is that of functor, i.e. morphism of categories; now, a functor is for us a kind of dictionary (it maps an object of the first category to an object of the second, and similarly for arrows) and, as we remarked above, functors are by no means sufficient to account for the majority of correspondences and dualities existing in mathematics. One needs a more flexible notion, and this is precisely what is achieved by the concept of a 'bridge' between different presentations of a topos.

The methodology of toposes as 'bridges' represents a structuralist way of doing mathematics in the sense that it is based on uncovering hidden structures (that is, structures that are for the most part invisible from the concrete perspective of the 'working mathematician', such as Morita equivalences and topos-theoretic invariants) and letting them guide the mathematical exploration; it is therefore 'upside-down' with respect to the more classical mathematical styles based on a preliminary study of 'concrete' structures and to the subsequent, ad hoc identification of suitable invariants.

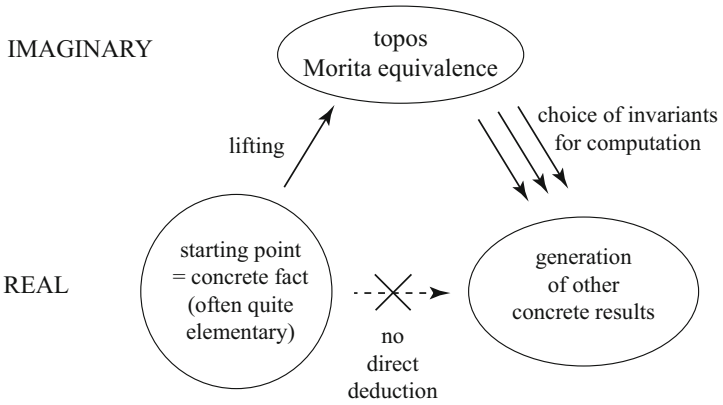
As far as the level of generality of the technique of topos-theoretic 'bridges' is concerned, this methodology is applicable to all those situations whose aspects that one wants to investigate can be encoded by means of suitable toposes and invariants on them. This certainly includes first-order mathematics (as formalized in terms of geometric logic) and a great amount of higher-order mathematics as well. Indeed, the possibility of considering Grothendieck toposes not just over the topos of sets but over an arbitrary Grothendieck topos allows one to construct classifying toposes for all those higher-order theories which can be formalized as (finite sequences of) relative geometric theories. Moreover, many mathematical objects of different kinds, including higher-order ones, can be used for presenting toposes; for example, the very notion of site is second-order.

9.12 The Duality Between 'Real' and 'Imaginary'

As already remarked in Sect. 9.10, the passage from a site (or a theory) to the associated topos can be regarded as a sort of 'completion' by the addition of 'imaginaries' (in the model-theoretic sense), which *materializes* the potential contained in the site (or theory). The duality between the (relatively) unstructured world of presentations of theories and the maximally structured world of toposes is of great relevance as, on the one hand, the 'simplicity' and concreteness of theories or sites makes it easy to manipulate them, while, on the other hand, computations are much easier in the 'imaginary' world of toposes thanks to their very rich internal structure and the fact that invariants live at this level.

The 'bridge' technique thus involves an ascent followed by a descent between two levels, the 'real' one of 'concrete' mathematics (represented by sites or other

objects presenting toposes) and the 'imaginary' one of toposes, which we can schematically represent as follows:



This 'jump' from the 'real' into the 'imaginary is indispensable, or at least highly useful, in many situations to reveal correspondences between 'concrete' mathematical contexts that would be hardly visible otherwise (cf. Sect. 9.11). Toposes thus act as sorts of 'universal translators', crucial for establishing the 'bridges' but disappearing in the final formulation of results.

References

Artin, M., A. Grothendieck, a,nd J.-L. Verdier. 1972. *Théorie des topos et cohomologie étale des schémas - (SGA4), Séminaire de Géométrie Algébrique du Bois-Marie, année 1963–64*. Lecture notes in mathematics, Vols. 269, 270 and 305. Springer.

Caramello, O. 2010. The unification of Mathematics via Topos Theory. arXiv:math.CT/1006.3930. Revised version to appear in the book "Logic in Question", Studies in Universal Logic, Springer (2021).

Caramello, O. 2014. Fraïssé's construction from a topos-theoretic perspective. *Logica Universalis* 8(2): 261–281.

Caramello, O. 2016a. Grothendieck toposes as unifying 'bridges' in mathematics. Mémoire d'habilitation à diriger des recherches, Université de Paris 7.

Caramello, O. 2016b. The theory of topos-theoretic 'bridges' – a conceptual introduction. Glass Bead Journal.

Caramello, O. 2016c. Topological Galois theory. *Advances in Mathematics* 291: 646–695.

Caramello, O. 2017. *Theories, sites, toposes: Relating and studying mathematical theories through topos-theoretic 'bridges'*. Oxford University Press.

Caramello, O. 2018. The idea of 'bridge' and its unifying role. Talk at TEDxLakeComo.

Caramello, O. 2020. La "notion unificatrice" de topos. To appear in the Proceedings volume of the *Lectures Grothendieckiennes*, École Normale Supérieure, Paris.

Krömer, R. 2007. *Tool and object: A history and philosophy of category theory*. Birkhauser.

Mac Lane, S. 1971. *Categories for the working mathematician*. Springer. Second edition: 1998.

Mac Lane, S., and I. Moerdijk. 1992. *Sheaves in geometry and logic: a first introduction to topos theory*. Springer.

Makkai, M., and G. Reyes. 1977. *First-order categorical logic*, volume 611 of *Lecture notes in mathematics* Springer.

Marquis, J.-P. 2010. *From a geometrical point of view: A study of the history and philosophy of category theory*. Springer.

Mazur, B. 2008. When is one thing equal to some other thing? In *Proof and other dilemmas: Mathematics and philosophy*. Mathematical Association of America.

Part III
Logics and Proofs

Chapter 10
Game of Grounds

Davide Catta and Antonio Piccolomini d'Aragona

Abstract In this paper, we propose to connect Prawitz's theory of grounds with Girard's Ludics. This connection is carried out on two levels. On a more philosophical one, we highlight some differences between Prawitz's and Girard's approaches, but we also argue that they share some basic ideas about proofs and deduction. On a more formal one, we sketch an indicative translation of Prawitz's theory grounds into Girard's Ludics relative to the implicational fragment of propositional intuitionistic logic. This may allow for a *dialogical* reading of Prawitz's ground-theoretic approach. Moreover, it becomes possible to provide a formal definition of a notion of *ground-candidate* introduced by Cozzo.

Keywords Grounding · Game · Dialogue · Proof · Pseudo-ground

10.1 Introduction

Dag Prawitz's *theory of grounds* and Jean-Yves Girard's *Ludics* are very recent semantics, shedding a new light upon some fundamental topics in contemporary mathematical logic.

Prawitz aims at explaining the compulsion exerted by proofs, in such a way that this epistemic power depends on the valid inferences of which a proof is built up. Girard, on the other hand, proposes a denotational semantics for Linear Logic, leading to a dialogical framework inspired by proof-search or game-theoretic approaches.

D. Catta (✉)
LIRMM – Montpellier University, Montpellier, France
e-mail: davide.catta@lirmm.fr

A. P. d'Aragona
Centre Gilles Gaston Granger, Aix-Marseille Univ, CNRS, Aix-en-Provence, France
e-mail: antonio.piccolomini-d-aragona@univ-amu.fr

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_10

The two theories have different targets and employ different formal means. Furthermore, they structurally differ on many points, often fundamental. Nevertheless, they also share some common philosophical standpoints. First of all, the idea that evidence depends on (the possession of) objects constructed through primitive, meaning-constitutive or meaning-conferring, operations. Secondly, the idea that a proof is not an evidence object, but an act through which an evidence object is obtained. Thus, our first aim in this paper is that of providing a philosophical comparison between Prawitz's theory of grounds and Girard's Ludics, by highlighting their similarities, as well as their differences (Sect. 10.4).

In the light of these philosophical similarities, one could wonder whether the two frameworks can be linked also in a more formal sense. As a second point of our work, we suggest, in a purely indicative way, how the link can be found, by outlining a translation from Prawitz's theory of grounds into Girard's Ludics (Sect. 10.5). The link could allow for a dialogical reading of Prawitz's account. In addition, it could permit a rigorous description of the notion of ground-candidate introduced by Cesare Cozzo to fix some weak points of an earlier formulation of Prawitz's theory of grounds.

Before establishing the aforementioned philosophical and – indicative – formal links, however, we need to introduce the basic concepts of Prawitz's theory of grounds and of Girard's Ludics. Thus, this will be our starting point (Sects. 10.2 and 10.3).

10.2 Theory of Grounds

Through the theory of grounds (ToG), Prawitz aims at explaining how and why a correct deductive argument might compel us to accept its conclusion if we have accepted its premises/assumptions. Prawitz's previous semantics in terms of valid arguments and proofs (SAP)[1] had exactly the same purpose, but it suffered from some problems. Clearly, the reasons that led Prawitz to ToG are not the topic of this paper. However, some of them must be mentioned, because they will be important when discussing the relationship between Prawitz's and Girard's respective philosophical standpoints.

10.2.1 From SAP to ToG

Prawitz's first semantic proposal[2] is centered on the notion of valid argument. An argument is a sequence of first-order formulas arranged in tree form, where

[1] See mainly Prawitz (1973, 1977).

[2] See mainly Prawitz (1973).

each node corresponds to an inference with arbitrary premises and conclusion. Validity of closed arguments – i.e. with no undischarged assumptions and unbound variables – is explained by saying that arguments ending with introductions, called canonical, are valid when their immediate sub-arguments are. Arguments ending with inferences in non-introductory form, called non-canonical, are instead valid when they can be reduced to a closed canonical valid argument. Open arguments are valid when all their closed instances are. Reductions employ constructive functions that transform trees ending with inferences in non-introductory form into trees where these inferences are removed. These functions are therefore a generalization of Prawitz's own reduction procedures for normalization in Gentzen's natural deduction systems.[3] In his later papers Prawitz proposes a similar approach, where the notion of valid argument in tree form is replaced by proofs understood as abstract objects built up of functions standing for valid inferences.[4] The definitions of notions such as closed canonical proof, closed non-canonical proof, and open proof, run basically like the corresponding ones in terms of valid arguments.

The question is now whether valid arguments and proofs, as defined in SAP, actually oblige us to accept their conclusion if we have accepted their assumptions. The epistemic power must stem from the valid inferences of which valid arguments and proofs are made up. Hence, what we need is an appropriate definition of the notion of valid inference. In SAP, this becomes the idea that an inference is valid if it gives rise to valid arguments and proofs when attached, respectively, to valid arguments and proofs. The overall framework, however, turns out to be unsatisfactory. One of the main problems stems from the interdependence of the concepts of valid inference and proof. The idea that proofs are chains of valid inferences seems to require a *local* notion of valid inference; in SAP, instead, the order of explanation is reversed, for valid inferences are defined in the *global* terms of the valid arguments and proofs they belong to. So, the compelling power cannot be explained by induction on the length of chains of valid inferences; since valid arguments and proofs may contain non-canonical steps, some valid inference may end a chain where a valid inference of the same kind, and of an equal or higher complexity may occur.[5]

Observe that the interdependence problem depends on the canonical/non-canonical distinction. So, a way out may be found by endorsing what is often called the *proof-objects/proof-acts* distinction introduced by Martin-Löf and developed by Sundholm.[6] The idea could be that proof-acts are chains of epistemic steps allowing to construct appropriate proof-objects. The former involve canonical or non-canonical moves, but we may require the latter to be always canonical, and

[3] See Gentzen (1934) and Prawitz (1965).

[4] See mainly Prawitz (1977).

[5] A similar point, although in a different context, is raised by Tranchini (2014a) and Usberti (2015).

[6] Martin-Löf (1984, 1986) and Sundholm (1998).

specified by simple induction.[7] Then, we may explain validity of inferences with respect to proof-objects, leaving proof-acts aside.

Clearly, valid inferences are exactly the steps that proof-acts are built up of. Hence, the above mentioned idea consists of explaining validity of inferences through the result that valid inferences yield – namely the proof-object they produce – rather than through how the result is obtained – namely the proof-act where they are used. This in turn means that what valid inferences generate is not only inferential structures, but abstract entities too. An inference cannot be the simple appending of a conclusion under some premises; also and above all, it has to be the application of an operation on proof-objects, with proof-objects as outputs in case of success.

As we shall see below, this strategy is exactly the one Prawitz follows with ToG.[8]

10.2.2 Grounds and their Language

The term ground is used by Prawitz to indicate "what a person needs to be in possession of in order that her judgment is to be justified or count as knowledge".[9] According to this view, one should conceive of "evidence states as states where the subject is in possession of certain objects".[10] But grounds also have to comply with a constructivist setup, whence they must be epistemic in nature. This means that they can be grasped according to the idea that "one finds something to be evident by performing a mental act".[11] So, such an informal picture can be further specified by spelling out what kind of constructive operations build grounds. Hence, grounds will be what we may call operational entities.

A language of grounds refers to a background language. Endorsing the so-called *formula-as-type conception*,[12] the formulas of the background language provide

[7] See also Tranchini (2014b).

[8] A more detailed reconstruction of the content of this Section is in Piccolomini d'Aragona (2017, 2019).

[9] Prawitz (2009), 187.

[10] Prawitz (2015), 88.

[11] Prawitz (2015), 88.

[12] See Howard (1980). The *formulas-as-types* conception is based on the idea that a formula should be understood as the class of its proofs, called its type. Thus, for example, a conjunction $A \wedge B$ corresponds to the cartesian product type $A \times B$, namely the class of pairs $\langle a, b\rangle$ with a object in the type A (proof of A) and b object in type B (proof of B); similarly, an implication $A \rightarrow B$ corresponds to the function space type $A \supset B$, namely, the class of functions $f(x^A)$ generating objects $f(g)$ in the type B when applied to objects g in the type A. The conception is consonant with the BHK interpretation of the meaning of the logical constants in terms of (canonical) proof-conditions of the formulas where such constants occur as main sign. In the case of an open formula $A(x_1, \ldots, x_n)$, the type is a function space, namely, the class of functions $f(x_1, \ldots, x_n)$ that produce objects in the type $A(k_1, \ldots, k_n)$ when applied to n individuals. Since the output type now depends on the input type, we may speak with Martin-Löf of *dependent types* (Martin-Löf,

types for the terms – and other components – of the language of grounds, while the latter are meant to denote grounds for asserting the former – the assertion sign being the fregean $\vdash$. ToG is mainly concerned with first-order logic. So, hereafter, the background language will be a first-order language. As one usually does in intuitionistic frameworks, we put

$$\neg A \stackrel{def}{=} A \to 0$$

– where 0 is an atomic constant for the absurd. Prawitz's main notions are relative to atomic bases. An *atomic base* for a language of grounds over a given background language will be identified by individual and relational constants of the background language, plus a (possibly empty) set of atomic rules over the background language. Atomic rules are conceived of by Prawitz in the framework of so-called Post-systems, i.e. recursive sets of rules

$$\frac{A_1, \ldots, A_n}{B}$$

over a first-order language L such that: (1) A_i and B are atomic, and $A_i \neq 0$ ($1 \leq i \leq n$); (2) if x occurs free in B, then there is $i \leq n$ such that x occurs free in A_i.[13] Once a set of atomic rules has been set down, atomic derivations can be specified in a standard inductive way.

We can outline a first example of language of grounds, labelled $\mathtt{C}$, where terms are built up of symbols that correspond only to Gentzen's introduction rules. Given a first-order language L and an atomic base $\mathfrak{B}$ over L with atomic system $\mathtt{S}$, the alphabet of $\mathtt{C}$ contains names c^A for atomic derivations of A in $\mathtt{S}$ with no undischarged assumptions and unbound variables, possibly indexed variables ξ^A typed on formulas A of L, typed operational symbols kI with $k = \wedge, \vee, \to, \forall, \exists$, and a typed operational symbol 0_A for A formula of L expressing the explosion principle. Typed terms are defined in a standard inductive way, e.g. in the case of $\vee$, $\to$, $\exists$ and 0 – we leave types of the operational symbols unspecified whenever possible –

- $T : A_i \in X \;\Rightarrow\; \vee I[A_i \vdash A_1 \vee A_2](T) : A_1 \vee A_2 \in X$ $[i = 1, 2]$
- $T : B \in X \;\Rightarrow\; \to I\xi^A(T) : A \to B \in X$ [the typed-variable after the symbol indicates that this variable is bound by the symbol]
- $T : A(t) \in X \;\Rightarrow\; \exists I[A(t) \vdash \exists x A(x)](T) : \exists x A(x) \in X$
- $T : 0 \in X \;\Rightarrow\; 0_A(T) : A \in X$

1984). One usually does the same with so-called hypothetical judgments $A_1, \ldots, A_n \vdash B$, where the input is given by objects in the types $A_1, \ldots, A_n$ and the output is an object in the type B.

[13] Sometimes, one also considers atomic rules that bind individual variables or assumptions, but this point is not essential for our purposes.

With C we can already state clauses fixing what counts as a ground for closed atomic formulas, for closed formulas with $\wedge$, $\vee$ and $\exists$ as main logical sign, and for 0:

(A) a ground over $\mathfrak{B}$ for atomic $\vdash A$ is any c^A
($\wedge$) a ground over $\mathfrak{B}$ for closed $\vdash A \wedge B$ is $\wedge I(T, U)$ where T denotes a ground over $\mathfrak{B}$ for $\vdash A$ and U denotes a ground over $\mathfrak{B}$ for $\vdash B$
($\vee$) a ground over $\mathfrak{B}$ for closed $\vdash A_1 \vee A_2$ is $\vee I[A_i \vdash A_1 \vee A_2](T)$ where T denotes a ground over $\mathfrak{B}$ for $\vdash A_i$ with $i = 1$ or $i = 2$
($\exists$) a ground over $\mathfrak{B}$ for closed $\vdash \exists x A(x)$ is $\exists I[A(t) \vdash \exists x A(x)](T)$ where T denotes a ground over $\mathfrak{B}$ for $\vdash A(t)$ for some t
(0) no T denotes a ground over $\mathfrak{B}$ for 0

On the contrary, the clauses for $\rightarrow$ and $\forall$ will involve total constructive functions of appropriate kind, taking individuals and/or grounds of given types as arguments, and producing grounds of given types as values. Like in the BHK-clauses,[14] the notion of total constructive function may be assumed as primitive:

($\rightarrow$) a ground over $\mathfrak{B}$ for closed $\vdash A \rightarrow B$ is $\rightarrow I\xi^A(T)$ where T denotes a ground over $\mathfrak{B}$ for $A \vdash B$, i.e. a constructive function over $\mathfrak{B}$ of type $A \vdash B$ [the typed variable after the symbol indicates that this variable is bound by the symbol]
($\forall$) a ground over $\mathfrak{B}$ for closed $\vdash \forall x A(x)$ is $\forall I x.(T)$ where T denotes a ground over $\mathfrak{B}$ for $\vdash A(x)$, i.e. a constructive function over $\mathfrak{B}$ of type $A(x)$ [the individual variable after the symbol indicates that this variable is bound by the symbol]

The clauses can be considered as a ground-theoretic determination of the meaning of the logical constants. Therefore, the operational symbols of C are primitive operations, and its terms may be qualified as canonical.

However, constructive functions have to be understood in an unrestricted way, and C is too weak to express all of them. Some terms denote some constructive functions, but not all functions are denoted – e.g. there is no term denoting a constructive function of type $A_1 \wedge A_2 \vdash A_i$ $(i = 1, 2)$. Thus, one must also bring in extensions of C – in fact, as we shall see in Sect. 10.2.3, because of Gödel's incompleteness no closed language of grounds permits to express all the grounds we need. The behaviour of an operation can be fixed through equations that show how to compute it on relevant values. As an example, consider the extension C^* of C whose alphabet contains new typed operational symbols kE with $k = \wedge, \vee, \rightarrow, \forall, \exists$, standing for Gentzen's eliminations, and additional, inductively defined terms, e.g. in the case of $\vee$, $\rightarrow$ and $\exists$ – we leave types of the operational symbols unspecified –

- $T : A \vee B, U : C, V : C \in X \Rightarrow \vee E\, \xi^A\, \xi^B.(T, U, V) : C \in X$ [the typed-variables after the symbol indicates that these variables are bound by the symbol.

[14] See for example Troelsta and Van-Dalen (1988).

Observe also that U and V are not necessarily of type $A \vdash C$ and $B \vdash C$ – these are just term-formation clauses. It is the equation below that permits us to interpret this operation semantically as one that yields a ground for $\vdash B$ when applied to grounds for $\vdash A \vee B$, $A \vdash C$ and $B \vdash C$ – think of the elimination rule for $\vee$ in Gentzen's natural deduction, and of the relative reduction as defined in Prawitz[15]]

- $T : A \to B, U : A \in X \Rightarrow \to E(T, U) : B \in X$
- $T : \exists x A(x), U : B \in X \Rightarrow \exists E\ x\ \xi^{A(x)}.(T, U) : B \in X$ [the typed- and individual variables after the symbol indicates that these variables are bound by the symbol. Observe also that U is not necessarily of type $A(x) \vdash B$ – these are just term-formation clauses. It is the equation below that permits us to interpret semantically this operation as one that yields a ground for $\vdash B$ when applied to grounds for $\vdash \exists x A(x)$ and $A(x) \vdash B$ – think of the elimination rule for $\exists$ in Gentzen's natural deduction, and of the relative reduction as defined in Prawitz[16]]

The equations are standard conversions, e.g. in the case of $\vee E$, $\to E$ and $\exists E$

- $\vee E\ \xi^{A_1}\ \xi^{A_2}.(\vee I[A_i \vdash A_1 \vee A_2](T), U_1(\xi^{A_1}), U_2(\xi^{A_2})) = U_i(T)$
- $\to E(\to I\xi^{A}(T(\xi^{A})), U) = T(U)$
- $\exists E\ x\ \xi^{A(x)}.(\exists I[A(t) \vdash \exists x A(x)](T), U(\xi^{A(x)})) = U(T)$

They show that the kE's capture new total constructive functions, e.g. in the case of $\vee E$ and $\exists E$ of types

1. $A \vee B, (A \vdash C), (B \vdash C) \vdash C$
2. $\exists x A(x), (A(x) \vdash B) \vdash B$

As a further example, consider the extension of C^* obtained by adding a typed operational symbol DS for disjunctive syllogism, with equations – we leave the type of DS unspecified –

- $DS(\vee I[B \vdash A \vee B](T), U) = T$
- $DS(\vee I[A \vdash A \vee B](T), U) = 0_B(\to E(T, U))$

At variance with C, the last two languages of grounds also contain non-canonical terms, that is, terms the outermost symbol of which is non-primitive.

Given a closed term T in some language of grounds over $\mathfrak{B}$, we now say that T denotes a ground over $\mathfrak{B}$ when it can be reduced to a term the outermost symbol of which is one of the operational symbols of C, that denotes a ground over $\mathfrak{B}$ according to one of the clauses $(\wedge)$ – $(\forall)$. Reduction employs the equations for the operational symbols of T, by replacing *definiendum* by *definiens*. For example

$$\to E(\to I\xi_1^{A\to A}(\vee\ \xi_2^{A\to A}\ \xi^0(\vee I[A \to A \vdash A \to A \vee 0](\xi_1^{A\to A}), \xi_2^{A\to A}, 0_{A\to A}(\xi^0))), \to I\xi^{A}(\xi^{A}))$$

[15] See Prawitz (1965).

[16] See Prawitz (1965).

by the equation for $\rightarrow E$ reduces to

$$\vee\, \xi_2^{A\rightarrow A}\, \xi^0(\vee I[A \rightarrow A \vdash A \rightarrow A \vee 0](\rightarrow I\xi^A(\xi^A)), \xi_2^{A\rightarrow A}, 0_{A\rightarrow A}(\xi^0))$$

which by the equation for $\vee E$ reduces to $\rightarrow I\xi^A(\xi^A)$. An open term denotes a constructive function over $\mathfrak{B}$ with domain $\mathfrak{D}$ and co-domain $\mathfrak{K}$ when all its closed instances denote a ground for $\mathfrak{K}$, where a closed instance is obtained by replacing individual variables with closed individual terms and typed variables with closed terms denoting grounds for the elements in $\mathfrak{D}$. So, if we replace $\rightarrow I\xi^A(\xi^A)$ with $\xi_3^{A\rightarrow A}$ in the example above, we obtain that the term we started from denotes a constructive function of type $A \rightarrow A \vdash A \rightarrow A$.[17]

10.2.3 Ground-Theoretic Validity

Prawitz affirms that "to *perform an inference* is, in addition to making an inferential transition, to apply an operation on the grounds that one considers oneself to have for the premises with the intention to get thereby a ground for the conclusion".[18] An inference will be valid over $\mathfrak{B}$ when the application of the corresponding operation to the alleged grounds over $\mathfrak{B}$ for the premises actually yields a ground over $\mathfrak{B}$ for the conclusion. The inference is logically valid if it remains valid over all the bases. A proof (over $\mathfrak{B}$) can be now defined as a finite chain of valid (over $\mathfrak{B}$) inferences. ToG respects the proof-as-chains intuition: the notion of proof is non-circularly stated in terms of a *local* notion of valid inference, in such a way that a proof involves only valid inferences. Crucially, "a proof of an assertion does not constitute a ground for the assertion but produces such a ground",[19] that is, ToG proofs are not objects but acts.

According to Prawitz's view, when an inferential agent carries a proof out, he/she comes into possession of a ground, not of a term. And grounds are, so to say, always canonical; terms describe how grounds are obtained – their being canonical or not depending on the kind of steps through which the possession is attained. Thus, when Gentzen's eliminations are linked to the non-primitive functional symbols of C^*, a performance of them on grounds for the premises will amount to nothing but a β-reduction. In more general cases, such as DS, we can in turn still resort to the old SAP idea of general procedures for obtaining canonical forms. Proofs can be conceived of as chains of applications of operations that, under relevant circumstances, coincide with computations on a sort of generalized non-normal forms.

[17] A better, but still partial development of the content of this Section is in Piccolomini d'Aragona (2018); a more detailed development is in Piccolomini d'Aragona (2019).

[18] Prawitz (2015), 94.

[19] Prawitz (2015), 93.

10.2.4 Cozzo's Ground-Candidates

Cozzo moved four objections to the old formulation of ToG presented in Prawitz's first papers on the subject.[20] Only the fourth one is of interest for us here. Prawitz's definition of an act of inference was such that only valid inferences were inferences whereas, as Cozzo argues, "it seems reasonable to say that the experience of necessity of thought also characterizes the transition from mistaken premises devoid of corresponding grounds".[21] In replying to Cozzo's objections, Prawitz relaxes the definition of inference act by allowing what he calls *alleged grounds*, i.e. "an [...] inference can err in two ways: the alleged grounds for the premisses may not be such grounds, or the operation may not produce a ground for the conclusion when applied to grounds for the premisses".[22] However, Usberti observes that "Prawitz's alleged grounds have nothing to do with ground-candidates: since no restriction is put on alleged grounds, they are entities of any kind. Now, an assertion based on an entity of any kind may be true or false, but it is difficult to see how it can be rational at all".[23] As Usberti observes, while Cozzo's first three objections might be overcome by Prawitz's adjustments, Cozzo's fourth objection remains unsolved.

As a remedy to his objections, Cozzo proposes an interesting notion of *ground-candidate*, "a mathematical representation of the results of epistemic acts underlying mistaken premises". A ground-candidate "can be a genuine ground or a *pseudo-ground*".[24] Can ground-candidates be characterized more precisely? As we shall see below, Girard's Ludics might suggest a promising account.

10.3 Ludics

Ludics was first proposed by Girard in his paper titled *Locus Solum: from the rules of logic to the logic of rules*.[25] Its aim is that of studying the notions of proposition and proof (of type and element of a type) and of reconstructing them from a more primitive notion of interaction.

As is well known, the cut-rule allows for an "interaction" between two proofs. Given a proof of A under hypotheses Γ, and a proof of B under hypotheses Δ, A, it yields a proof of B under hypotheses Γ, Δ. Gentzen's *Haupsatz* establishes that a derivation π of A under Γ can be always "normalized" to a derivation π' of A under no more hypotheses than those in Γ, and where no cut-rule is employed. In

[20] See Prawitz (2009, 2012).
[21] Cozzo (2015), 114.
[22] Prawitz (2015), 95.
[23] Usberti (2017), 525.
[24] Cozzo (2015), 114.
[25] See Girard (2001).

the light of the Curry-Howard correspondence, normalization can be understood as the evaluation of a program to its normal form.

Usually, the cut-rule is conceived as a deduction rule over a previously defined formal language. Once language and rules have been given, one can study the properties of cut-elimination. In order to attribute a computational meaning to proofs, a deterministic cut-elimination is required, i.e. given an instance of the cut-rule, there is only one way of eliminating it.

Ludics starts from an entirely opposite point of view. Given a deterministic cut-elimination defined on *sui generis* computational objects, it seeks whether it is possible to recover formulas and rules. The objects of Ludics are an abstract counterpart of derivations in a well-behaved and complete fragment of multiplicative-additive Linear Logic. Such objects are built up of rules that ensure a deterministic cut-elimination on arbitrary numerical addresses. The latter represent the location that a formula may occupy in a derivation. This innovative approach, of which we shall say more below, is mainly inspired by a polarity phenomenon that we deal with in the next Section.

10.3.1 Polarity

Linear connectives are normally divided into two classes: of positive polarity – i.e. $\otimes$, multiplicative conjunction, $\oplus$, additive disjunction, 1 multiplicative truth and 0 additive false – and of negative polarity – i.e. &, additive conjunction, $\parr$, multiplicative disjunction, $\top$, additive truth and $\bot$ multiplicative false. A formula is said to be negative (resp. positive) iff its main connective is negative (resp. positive). A linear connective $\star$ is said to be reversible iff, for each proof π of a sequent Γ, A, for A with main connective $\star$, there is a cut-free proof π' of Γ, A, the last rule of which introduces $\star$.[26] The two notions are connected, in that negative connectives are reversible, whilst the positive ones are not.

Polarity and reversibility are very important for proof-search. Suppose we are looking for a proof of a sequent $\Gamma = \Gamma', A$, i.e. we are trying to prove Γ by selecting a formula A in Γ and then by applying an inference on it. If A is negative, the reversibility of its main connective ensures the existence of a cut-free proof of Γ', A, the last rule of which introduces $\star$. And this means that if, e.g., $A = B \parr C$, then we have only one bottom-up application of the rule, so that Γ', B, C is provable too. Thus, selecting a negative formula in a sequent allows for an automatic proof-search procedure. But this strategy cannot be applied uniformly, because of the non-reversibility of positive connectives.

This notwithstanding, proof-search procedures can still be improved thanks to the following algorithm introduced by Andreoli.[27] Given a sequent Γ: (1) if it contains

[26] See Laurent (2002) for a detailed discussion of the phenomenon of polarity in Linear Logic.

[27] See Andreoli (1992).

negative formulas, focus on them and apply negative rules upwards until they are all decomposed, and there are no more negative formulas; (2) when you have finished, choose a positive formula randomly, and start decomposition by applying positive rules upwards, focusing at each step on its positive sub-formulas until you find negative formulas. Then, Andreoli proves that a sequent is derivable in Linear Logic, iff there is a focusing derivation of it, i.e. a derivation that can be built by applying his algorithm. It follows that one can restrict oneself to a set of derivations the elements of which consist of alternations of positive and negative steps only; positive steps cluster the positive branch of the algorithm – labelled (2) above – whereas negative steps cluster the negative branch of the algorithm – labelled (1) above. As an example of proof obtained through Andreoli's algorithm, consider the following one:

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\vdash A^{\perp}, A} \quad \dfrac{}{\vdash B^{\perp}, B}}{\vdash A, B, (\mathbf{A}^{\perp} \otimes \mathbf{B}^{\perp})}}{\vdash A, B, (\mathbf{A}^{\perp} \otimes \mathbf{B}^{\perp}) \oplus (\mathbf{A}^{\perp} \otimes \mathbf{C}^{\perp})} \quad \dfrac{\dfrac{\dfrac{}{\vdash A^{\perp}, A} \quad \dfrac{}{\vdash C^{\perp}, C}}{\vdash A, C, (\mathbf{A}^{\perp} \otimes \mathbf{C}^{\perp})}}{\vdash A, C, (\mathbf{A}^{\perp} \otimes \mathbf{B}^{\perp}) \oplus (\mathbf{A}^{\perp} \otimes \mathbf{C}^{\perp})}}{\dfrac{\vdash A, (\mathbf{B} \,\&\, \mathbf{C}), (A^{\perp} \otimes B^{\perp}) \oplus (A^{\perp} \otimes C^{\perp})}{\vdash \mathbf{A} \parr (\mathbf{B} \,\&\, \mathbf{C}),\ (A^{\perp} \otimes B^{\perp}) \oplus (A^{\perp} \otimes C^{\perp})}}$$

Here, the bold formulas are those chosen for the step-wise proof-search procedure. The proof can be transformed into one where the two negative steps of the first sub-proof are merged into a single negative step, and where the same happens for the corresponding positive steps, i.e.

$$\dfrac{\dfrac{\dfrac{}{\vdash A^{\perp}, A} \quad \dfrac{}{\vdash B^{\perp}, B}}{\vdash A, B, (\mathbf{A}^{\perp} \otimes \mathbf{B}^{\perp}) \oplus (\mathbf{A}^{\perp} \otimes \mathbf{C}^{\perp})} \quad \dfrac{\dfrac{}{\vdash A^{\perp}, A} \quad \dfrac{}{\vdash C^{\perp}, C}}{\vdash A, C, (\mathbf{A}^{\perp} \otimes \mathbf{B}^{\perp}) \oplus (\mathbf{A}^{\perp} \otimes \mathbf{C}^{\perp})}}{\vdash \mathbf{A} \parr (\mathbf{B} \,\&\, \mathbf{C}), (A^{\perp} \otimes B^{\perp}) \oplus (A^{\perp} \otimes C^{\perp})}$$

In this way, polarization permits to consider *generalized connectives*, obtained by "merging" negative connectives with negative connectives, and positive connectives with positive connectives. In our example, the negative generalized connective is $- \parr (- \,\&\, -)$ and the positive generalized connective is $(- \otimes -) \oplus (- \otimes -)$. In addition, the sequents of the calculus end up having the form Γ, Δ, with Γ containing only positive formulas and Δ containing at most one negative formula. Although an arbitrary number of such connectives exists, all of them can be decomposed according to the same schemes of rules as follows – we use P as a variable for positive formulas and N as a variable for negative formulas:

$$\dfrac{\vdash N_{i_1}, \Gamma_1 \quad \cdots \quad \vdash N_{i_{n_i}}, \Gamma_n}{\vdash (N_{1_1} \otimes \cdots \otimes N_{1_{n_1}}) \oplus \cdots \oplus (N_{p_1} \otimes \cdots \otimes N_{p_{n_p}}), \Gamma_1, \ldots \Gamma_n} \text{ positive rule}$$

$$\dfrac{\vdash P_{1_1} \cdots P_{1_n}, \Gamma \quad \cdots \quad \vdash P_{1_{p_1}} \cdots P_{n_{p_n}}, \Gamma}{\vdash (P_{1_1} \parr \cdots \parr P_{1_n}) \,\&\, \cdots \,\&\, (P_{1_{p_1}} \parr \cdots \parr P_{1_{p_n}}), \Gamma} \text{ negative rule}$$

Thanks to the fact that De Morgan dualities hold for linear logic, and that the linear negation of a negative formula is a positive formula and vice versa, we can restrict our language to formulas composed by $\otimes, \oplus$ and write the schemes of rules above as follows:

$$\frac{P_{i_1} \vdash, \Gamma_1 \quad \cdots \quad P_{i_{n_i}} \vdash \Gamma_n}{\vdash (P_{1_1}^{\perp} \otimes \cdots \otimes P_{1_{n_1}}^{\perp}) \oplus \cdots \oplus (P_{p_1}^{\perp} \otimes \cdots \otimes P_{p_{n_p}}^{\perp}), \Gamma_1, \ldots \Gamma_n} \text{ positive rule}$$

$$\frac{\vdash P_{1_1} \cdots P_{1_n}, \Gamma \quad \cdots \quad \vdash P_{1_{p_1}} \cdots P_{n_{p_n}}, \Gamma}{(P_{1_1}^{\perp} \otimes \cdots \otimes P_{1_n}^{\perp}) \oplus \cdots \oplus (P_{1_{p_1}}^{\perp} \otimes \cdots \otimes P_{1_{p_n}}^{\perp}) \vdash \Gamma} \text{ negative rule}$$

10.3.2 From Polarity to Games

Observe that each derivation in a calculus based upon the clustered rules of the previous Section is a tree in which positive rules are followed by negative rules and vice-versa – except the axiom rule. This allows us to consider "games" on a sequent.

Let us say that a *move* is a couple $(F, \{F_1, \ldots, F_n\})$, where F is a positive (negative) formula, called *focus*, and $F_1, \ldots, F_n$ are some negative (positive) subformulas of F, called *choices*. A move is of *positive polarity* if its focus is a positive formula, and of *negative polarity* otherwise. So, a *game* is a non-empty sequence of moves such that: (1) two consecutive moves have opposite polarity; (2) the focus of a negative rule, except possibly the first, is one of the formulas in the choices of the positive move that immediately precedes it; (3) two distinct moves have distinct focuses. We finally define a strategy on $\Gamma \vdash \Delta$ to be a prefix-closed set of games on the same sequent. It is easily seen that the polarized clustered derivation in the previous Section can be considered as the set built up of the two games

$$\mathfrak{G}_1 = (A \parr (B \,\&\, C), \{A, B\}), ((A^{\perp} \otimes B^{\perp}) \oplus (A^{\perp} \otimes C^{\perp}), \{A^{\perp}, B^{\perp}\})$$

$$\mathfrak{G}_2 = (A \parr (B \,\&\, C), \{A, C\}), ((A^{\perp} \otimes B^{\perp}) \oplus (A^{\perp} \otimes C^{\perp}), \{A^{\perp}, C^{\perp}\})$$

The vice-versa holds too, i.e. the strategy can be converted into the proof above.

Polarity is therefore useful to understand one of the main ideas that, as we shall see, inspire Girard's Ludics: proofs are strategies over a particular type of game. For this to make sense, though, we take a final step. The cut-rule

$$\frac{\vdash P \qquad \vdash P^{\perp}}{\vdash}$$

can be interpreted as indicating an interplay between a strategy π for P and an opposed strategy π' for $P^{\perp}$. Clearly, if one between P and $P^{\perp}$ is provable, the

other cannot be so. Thus, if we want that every strategy "corresponds" to a proof in a polarized formal system as above, we need to admit proofs of both a proposition and its negation. To this end, we may introduce a new rule – call it *daimon* – that permits to prove any possible sequent:

$$\frac{}{\vdash \Gamma}\ \dagger$$

10.3.3 Ludics Defined

Ludics aims at overcoming the distinction between syntax (a formal system) and semantics (its interpretation). As is well-known, completeness implies that, for every A, either there is a proof of A or a counter-model of A, i.e. a model of $\neg A$. This can be understood through a dialogical metaphor; in a two-persons debate, where one of the speakers tries to construct a proof of A and the other tries to construct a counter-model of A, one and only one of the debaters can win. However, models and proofs are distinct entities and, in particular, there is no interaction between a proof of A and a model of $\neg A$. Girard's idea is that of overcoming the syntax-semantic distinction by interpreting proofs with proofs, and counter-models with refutations. The properties of the attempted proofs of A are verified by testing them, by means of a cut-elimination procedure, against attempted proofs of $\neg A$. Proofs should be understood as *proof-search procedures* in the setup of the polarized system sketched in the two previous Sections. Given a certain conclusion, we try to guess the inference rule used to obtain it. If no rule is applied, we close the proof-search by "giving up", i.e. by a rule that encodes the information "I do not know how to keep on proving the conclusion, therefore I simply assert it". This rule is what permits to have attempted proofs of both A and $\neg A$.

A Ludics derivation consists of rules extracted from those occurring in the above-told clustered derivations. Formulas are replaced by numerical addresses standing for the positions they may occupy during proof-search. In order to make this more precise, we now give a quick sketch of formalized Ludics.

We say that an *address* is a string of natural numbers $i_1 \cdots i_n$. Two addresses are said to be *disjoint* if none of them is a prefix of the other. A *ramification* is a finite set of natural numbers. Given an address ξ and a natural number i, with $\xi\ i$ we indicate the address obtained by putting i at the end of ξ. Given an address ξ and a ramification I, with $\xi \star I$ we indicate the set of the addresses $\{\xi\ i \mid \text{for every } i \in I\}$. A *pitchfork* is an expression $\Gamma \vdash \Delta$ where Γ, Δ are finite sets of pairwise disjoint addresses such that Γ contains at most one address. Addresses in Γ are called *negative*, while those in Δ are called *positive*. A *design* is a tree made of pitchforks, the last pitchfork being the *base*, while the others are built through the *rules*:
daimon-rule [already introduced in the previous Section]

$$\frac{}{\vdash \Gamma}\ \dagger$$

Positive rule
Let I be a ramification and, for every $i \in I$, let the Γ_i be pairwise disjoint and included in Γ. For every $i \in I$, the rule (finite) is

$$\frac{\cdots\ \xi \star i \vdash \Gamma_i\ \cdots}{\vdash \Gamma, \xi}\ (\vdash \xi, I)$$

Negative rule
Let N be a set of ramifications and, for every $I \in N$, let Γ_I be included in Γ. For every $I \in N$, the rule (possibly infinite) is

$$\frac{\cdots \vdash \Gamma_I, \xi \star i\ \cdots}{\xi \vdash \Gamma}\ (\xi \vdash N)$$

The design – not to be confused with the only rule applied in it –

$$\frac{}{\vdash \Gamma}\ \dagger$$

is called *Daimon*. The only design not built by employing the rules above – called *Fid* and responding to the idea of a positive conclusion with no rules above – is

$$\frac{}{\vdash \Gamma}\ \Omega$$

A *cut* is given by an address that occurs once at the right of $\vdash$ (positive polarity) in the base of a design $\mathfrak{D}$, and once at the left of $\vdash$ (negative polarity) in the base of a design $\mathfrak{D}'$. Thus, a *cut-net* is a finite non-empty set of designs such that: (1) the addresses in the bases are pairwise disjoint or equal; (2) every address occurs in at most two bases and, if it occurs in two bases, it occurs both positively and negatively; (3) the graph with vertices the bases and edges the cuts is connected and acyclic. The *principal design* of a cut-net $\mathfrak{R}$ is the only design $\mathfrak{D} \in \mathfrak{R}$ with base $\Gamma \vdash \Delta$ such that Γ is not a cut. The *base* of a cut-net are the uncut addresses of the cut-net.

We now proceed to an informal description of the process of normalization/interaction on *closed* cut-nets, i.e. cut-nets the base of which is empty.[28] Given such a cut-net, the cut propagates over all the immediate sub-addresses as long as the action anchored on the positive pitchfork containing the cut corresponds to one of the actions anchored on the negative one. The process terminates when either the positive action anchored on the positive cut-fork is the daimon-rule, in which case we obtain a design with the same base as the starting cut-net, or no negative action corresponds to the positive one. In the latter case, the process is said to diverge. When normalization/interaction between two designs $\mathfrak{D}$ and $\mathfrak{D}'$ terminates and does not diverge, it ends up in Daimon, and $\mathfrak{D}$ and $\mathfrak{D}'$ are said to be orthogonal – indicated with $\mathfrak{D} \perp \mathfrak{D}'$. We give an example of terminating normalization on a closed cut-net composed of two designs. Bold addresses are those through which the normalization/interaction procedures propagates.

[28] For a complete and formal definition, also on the open case, see Girard (2001).

$$\dfrac{\dfrac{\dfrac{}{\vdash \boldsymbol{\xi 11}}\,\dagger \quad \dfrac{}{\vdash \xi 13}\,\emptyset}{\boldsymbol{\xi 1} \vdash}}{\vdash \boldsymbol{\xi}} \qquad \dfrac{\dfrac{\dfrac{\vdots}{\boldsymbol{\xi 11} \vdash}}{\vdash \boldsymbol{\xi 1}} \quad \dfrac{\vdots}{\vdash \xi 2}}{\boldsymbol{\xi} \vdash}$$

$$\dfrac{\dfrac{\dfrac{}{\vdash \boldsymbol{\xi 11}}\,\dagger \quad \dfrac{}{\vdash \xi 13}\,\emptyset}{\boldsymbol{\xi 1} \vdash} \qquad \dfrac{\dfrac{\vdots}{\boldsymbol{\xi 11} \vdash}}{\vdash \boldsymbol{\xi 1}}$$

$$\dfrac{\dfrac{}{\vdash \boldsymbol{\xi 11}}\,\dagger \qquad \dfrac{\vdots}{\boldsymbol{\xi 11} \vdash}$$

$$\dfrac{}{\vdash}\,\dagger$$

Generalizing, we say that the *ortoghonal* of a set E of designs on the same base $\Gamma \vdash \Delta$ – written $E^{\perp}$ – is the set

$$\{\mathfrak{D} \mid \text{for all } \mathfrak{D}' \in E, \mathfrak{D} \perp \mathfrak{D}'\}$$

For later purposes, we now need only four additional definitions. First of all, we say that a set of designs E is a behaviour when $E = E^{\perp\perp}$. We give two simple examples of behaviours. Take any design of the form

$$\dfrac{}{\vdash \xi}\,\emptyset$$

this type of design is called *atomic bomb*, and take E to be the singleton set containing the atomic bomb. The set $E^{\perp}$ only contains

$$\dfrac{\dfrac{}{\vdash}\,\dagger}{\xi \vdash}\,\emptyset$$

This last design is also orthogonal to the Daimon

$$\dfrac{}{\vdash \xi}\,\dagger$$

Thus the set $E^{\perp\perp}$ will contain the atomic bomb and the Daimon – this behaviour represents the constant 1. Consider now the following design, called *skunk*.

$$\dfrac{}{\xi \vdash}\,\emptyset$$

Its only orthogonal will be the Daimon. Call dai the set that only contains the Daimon. $dai^{\perp}$ will contain the skunk and any design of the form

$$\dfrac{\vdots}{\xi \vdash}\,N$$

where N is a subset of the finite part of the power-set of the natural numbers. We denote this last behaviour by $\top$.

Before giving the second definition, observe that in the example of normalization above not all the addresses of the two designs that form the cut-net are explored by the normalization procedure, which means that in general, given a design $\mathfrak{D}$, only a sub-design $\mathfrak{D}'$ of $\mathfrak{D}$ is used to test a counter-design of opposite base. To get a clearer idea, consider the behaviour $\top$ defined above. The orthogonal of $\top$ is the behaviour dai. Given any design $\mathfrak{D}$ in $\top$, normalization between $\mathfrak{D}$ and the Daimon will not explore the addresses open by negative rules. So, the only design that is used in the interaction between an element of $\top$ and † is the skunk. This leads us to the following: given a design $\mathfrak{D}$ and a behaviour G, we call *incarnation* of $\mathfrak{D}$ in G – indicated with $|\mathfrak{D}|_G$ – the smallest sub-design $\mathfrak{D}'$ of $\mathfrak{D}$ which is still in G. As a third definition, we say that $\mathfrak{D}$ is *material* in a behaviour G when $\mathfrak{D} = |\mathfrak{D}|_G$. The skunk in $\top$, as well as the two designs in 1 are material. Finally, given a behaviour G, we call *incarnation* of G – indicated with $|G|$ – the set

$$\{\mathfrak{D} \mid \mathfrak{D} \in G \text{ and } \mathfrak{D} = |\mathfrak{D}|_G\}$$

To conclude, two points must be underlined. First of all, † stands for paralogism; it captures the idea of abandoning the dialogue or game on a position that one is not able to justify further. When interaction yields †, one of the two designs is a locally winning strategy; though, the same strategy may lose in other contexts. A globally winning strategy is instead one that can never be defeated; a design can thus be understood as a proof when the interaction with any of its ortoghonals gives †. It follows that designs are not necessarily proofs – they may be nothing but attempted proofs. Secondly, designs are cut-free; cuts only occur when interaction is defined. The idea is that a type is a set of cut-free attempted proofs; more precisely, a type amounts to a set of cut-free paraproofs that behave in the same way in interactions with orthogonal designs. More precisely, in Ludics a type is nothing but a behaviour, as the latter notion has been defined above.

10.4 Differences and Similarities

We now propose a philosophical comparison between ToG and Ludics. Although some deep differences may be detected, we shall argue that these two theories share equally deep tenets, somehow inspiring the general framework they provide.

10.4.1 Differences: Order, Types and Bidirectionalism

The first and most striking difference between ToG and Ludics concerns the logic which they aim at interpreting. Girard's Ludics can be considered as an

interpretation of second-order multiplicative-additive Linear Logic with weakening. Every formula A is interpreted as a behaviour, and this permits to prove that: (1) if π is a proof of A, then there exists a †-free, material design $\mathfrak{D}$ in the behaviour associated to A such that $\mathfrak{D}$ is the interpretation of π; (2) if $\mathfrak{D}$ is a material and †-free design in a behaviour A, then $\mathfrak{D}$ is the interpretation of a proof π of A. Prawitz's ToG, instead, mainly aims at interpreting first-order intuitionistic logic. As is well-known, a Linear interpretation of first-order intuitionistic logic requires the modality operator !, allowing for contraction. So, for example, the intuitionistic $A \to B$ can be defined as $(!A) \multimap B$. In the version of Ludics we have been referring so far, no interpretation of the type $!A$ is given. However, such interpretation *can* be given, for example in *C-Ludics*,[29] or in *Ludics with repetitions*.[30] A difference between C-Ludics and Ludics with repetitions is that the designs of C-Ludics may contain cuts, whereas those in Ludics with repetitions – as already in Girard's original formulation – are always cut-free. Since the cut-free character of designs will be one of the common points we will highlight between ToG and Ludics, the comparison we propose should be thought as referring to Ludics with repetitions, rather than to C-Ludics.

This first kind of difference also concerns the order on which Prawitz and Girard reason. Prawitz's grounds and terms of a language of grounds are thought of as referring to a *first-order* background language. The latter provides types for either classifying abstract objects, or labelling syntactical expressions that denote such objects. Higher-order theories of grounds may be of interest, but they would face the same difficulties as those met in any other constructivist approach. In fact, it is well-known that nth-order logics for $n \geq 2$ imply a loss of what Dummett called molecularity on introduction rules.[31] So, a ground for a second-order existential $\vdash \exists X \alpha(X)$ should be defined through a primitive operation, say $\exists^2 I$, by requiring that $\exists^2 I(\mathfrak{U}, T, X)$ denotes a ground for $\vdash \exists X \alpha(X)$ iff T denotes a ground for $\vdash \alpha(\mathfrak{U})$; but $\mathfrak{U}$ may contain $\exists X \alpha(X)$ as a sub-formula, and the definition could be no longer compositional. In addition, higher-order constructivist setups might suffer from impredicativity and paradoxical phenomena.[32] On the contrary, Ludics is a *second-order* theory, and this would appear as an unsurmontable obstacle between Prawitz's and Girard's approaches.

Recent works, however, have shown that Ludics is suitable for first-order quantification,[33] or for first-order Martin-Löf dependent types.[34] Moreover, the propositional level is much less problematic – although, as we will see, not entirely unproblematic. The difference concerning logics is therefore undoubtedly important, but not as worrying as it may appear at a first glance.

[29] See Terui (2011).

[30] See Faggian and Basaldella (2011).

[31] See Dummett (1993) and Cozzo (1994).

[32] See Pistone (2018).

[33] See Fleury and Quatrini (2004).

[34] See Sironi (2014a,b).

The second difference is what we may call the *typed vs untyped* distinction. Inspired by the Curry-Howard isomorphism, ToG adopts the so-called *formulas-as-types* conception, in the light of which, as already remarked, a background language provides *types* for grounds and terms. Conversely, Ludics is fully *untyped*. Nonetheless, Ludics aims at recovering types out of a more primitive notion of interaction. In fact, we *do have* types herein; they are reconstructed as sets of designs respecting certain constraints, namely, sets of designs equal to their bi-orthogonal, namely, behaviours. By attributing a pivotal role to the notion of interaction, Ludics focuses on typability, rather than on typedness; metaphorically, it seeks the conditions under which a dialogue obtains, as well as the conditions under which a dialogue can be said to employ logical means. So, while in ToG types are primitive sets of proofs, in Ludics they are non-primitive sets of well-behaving designs. And since a design can be understood as a paraproof, proofs will only constitute the *subset* of a type – i.e. the set of the designs of a behaviour that win in every possible interaction.

The third and deeper difference between Prawitz's and Girard's approaches is finally the following; while ToG is clearly a verificationist semantics, in that it bases the explanation of meaning on primitive operations that mirror Gentzen's introduction rules, Ludics seems to be more akin to bidirectionalism.

Verificationism can be roughly described as the idea of explaining meaning in terms of the *conditions* for asserting correctly; its dual is *pragmatism*, according to which meaning should be fixed in terms of the correct *consequences* of an assertion. Gentzen's natural deduction provides a paradigmatic picture for both verificationism and pragmatism; the former is centred on introduction rules, with respect to which elimination rules have to be justified, whereas the latter proceeds the other way round.

Bidirectionalism is instead the standpoint according to which meaning must be explained in terms of *two* primitive notions; the conditions for asserting correctly, and the conditions under which one can safely modify the assumption of an assertion.[35] Not surprisingly, bidirectionalism fits better with Gentzen's sequent calculus that comes with two kinds of introduction rules – right introductions and left introductions.

As remarked by Schroeder-Heister, left introductions for the logical constant k can be understood as generalized elimination rules where a major premise with main sign k occurs as an assumption, and replaces assumptions of lower complexity, which may in turn be discharged. Now, it can be shown that negative "clustered" rules, as discussed in Sect. 10.3.1, fit with generalized elimination rules of this kind. To verify why Ludics is more akin to bidirectionalism, it is sufficient to recall that its rules abstract from the "clustered" ones; *positive* Ludics rules mirror right

[35] For the Linear Logic framework, see Zeilberger (2008); for the natural deduction framework, see Schroeder-Heister (2009); an interesting discussion of bidirectionalism in connection with grounding can be found in Francez (2015) – Francez's grounding is however entirely different from Prawitz's one.

introductions, whilst negative Ludics rules mirror left introductions, i.e. generalized eliminations. Crucially, both positive and negative rules are *primitive*, relative to the determination of types.

10.4.2 Similarities: Objects/Acts and Computation

Despite the more or less relevant differences, it seems to us that ToG and Ludics share at least two basic ideas. These tenets somehow inspire, from a philosophical as well as from a formal standpoint, the overall framework which Prawitz and Girard envisage for a rigorous reconstruction of deduction. Both such ideas can be understood through the lens of the aforementioned distinction between *proof-objects* and *proof-acts*.

As previously written, the distinction is explicitly at play in ToG; on the one hand, we have grounds – what one is in possession of when justified – and, on the other, we have proofs – acts by means of which grounds are obtained. These two sides, which we may call respectively *abstract* and *operational*, interact via the terms of the languages of grounds; a term denotes a ground, and codes a proof delivering the denoted ground.

Also Ludics involves an abstract and an operational side. Here, the objects are attempted proofs – designs – while the dynamics of deduction is represented by interaction – cut-elimination in cut-nets. Undoubtedly, Girard proposes a very peculiar standpoint about the act of (para)proving, i.e. a dialogical or game-theoretic one. Nevertheless, there is still the idea that a (para)proof-act produces a (para)proof-object; interaction looks for the (locally or globally) winning design and, if converging, it results in a normal form representing a strategy that the (locally or globally) winner may endorse for being (locally or globally) justified.

Now, both in ToG and in Ludics, the abstract/operational articulation relies upon what seems to be the same programmatic idea – being the first similarity we are going to point out. Prawitz's objects are, as we have seen, always *canonical*, since grounds are always obtained by applying primitive operations; ToG acts, instead, may be *canonical or not*, depending on how the ground they yield is obtained. Likewise, Girard's objects, i.e. designs, are always *cut-free*; *cuts* only occur in cut-nets, i.e. in interaction.

Thus, to Prawitz's abstract level, inhabited by canonical objects, corresponds Girard's abstract level, inhabited by normal designs. Indeed, canonicity can be looked at as a semantic generalization of normal form, where derivations are taken not as elements generated by a fixed formal system, but as broad structures through which one defines notions such as validity and consequence.[36] To Prawitz's operational level, where we do have a distinction between canonical and non-

[36] See Schroeder-Heister (2006), about the so-called *fundamental corollary* of Prawitz's normalization theory.

canonical cases, corresponds Girard's operational level, where we do have cuts. Once again, we have a semantic generalization of cut-elimination; to justify a non-canonical step, one requires harmony with respect to the canonical cases, and this is done by showing how maximal peaks that the non-canonical step would create may be appropriately dropped out.

Reasonably, one could now wonder where such a symmetry stems from. To answer this question, we have to turn to the second similarity between ToG and Ludics. Recall that Prawitz's notion of ground accounts for evidence. But justification takes root in meaning. Hence, grounds are built up of *primitive* and *meaning-constitutive* operations. Within Ludics, the parallel idea is that (tentative) evidence, represented by designs, depends on some *primitive* rules, the interaction of which in cuts allows for *determination of meaning* of definable types. Non-canonicity and cuts appear as soon as we focus, not on what evidence is, but on how evidence is acquired. This may be because for both Prawitz and Girard a proof is an act of a very special kind: it is a *computation* that removes non-primitive elements.

To be more precise, we know that Prawitz takes inferences as applications of operations on grounds. Given the way operations on grounds are understood, i.e. as fixed through equations that show how to compute non-canonical terms, we have pointed out that, under relevant circumstances, inferences could be understood as *generalized reduction steps*. Analogously, in Girard's framework, interaction is a process through which a (locally or globally) winning design may be found at the end of a dialogue or game between opposite positions. The act of proving may be therefore understood as exactly this proof-search, proceeding via *cut-elimination*.

However, the parallelism between the abstract sides of Prawitz's and Girard's theories undergoes some restrictions. In fact, the claim that to Prawitz's objects – the grounds – correspond Girard's object – the designs – must be accompanied by the observation that designs are not objects on the same level or of the same nature as grounds. A ground reifies an evidence state for a *specific* judgement or assertion; as such, it is of a *specific* type, according to the proposition or sentence involved in the judgment or assertion for which it is a ground. On the other hand, a design reifies the *moves* in a proof-game, or in proof-search, independently from judgements or assertions at issue in the interplay; as already remarked, Ludics is untyped precisely because it aims at reconstructing the pure *dynamics* of giving and asking for reasons, and at recovering typed logical strategies out of a more basic notion of interaction. To use a slogan, we could say that, while Prawitz's objects are objects in the "Fregean" sense of being saturated entities obtained by filling unsaturated entities, Girard's objects are "interactional" objects.

On the other hand, Girard's designs come conceptually closer to Prawitz's grounds when they are looked at within the global context of a behaviour, i.e. when they share stable interactional properties with other designs. Once a class of strategies has been, so to say, closed under cut-elimination, we are allowed to speak of a type. Typing is thus obtained by abstracting from overall properties that objects of the same type are expected to show when used in (attempted) deduction. This is also the linchpin of our indicative proposal for formally linking Prawitz's ToG and Girard's Ludics.

10.5 Grounds in Ludics

After the philosophical comparison between ToG and Ludics, we seek whether these two theories can also be compared on a more formal level. We remark, however, that what follows is just an introductory framework, requiring further work and refinement.

10.5.1 A Translation Proposal: The Implicational Fragment

Before starting, we pay our due to Sironi.[37] When proposing an understanding of Prawitz's grounds in Girard's framework, we will largely rely upon her embedding of Martin-Löf's type theory in Ludics.

By limiting ourselves to the implicational fragment of intuitionistic logic, we show in a purely indicative way how Prawitz's grounds for formulas of the kind $A \to B$ could be translated onto Girard's framework as designs respecting certain constraints – we discuss only in the conclusive remarks to what extent the translation may apply to the other intuitionistic first-order constants.

The leading idea of our indicative mapping is that of undertaking Girard's own line of thought, and looking at a type A as a behaviour G^A. A ground for $\vdash A$ is to be a †-free element of $|G^A|$, the incarnation of G^A. Finally, a cut-net $\mathfrak{R}$ between two appropriate †-free elements of incarnations, and such that $[[\mathfrak{R}]]$ is a †-free element of $|G^A|$, will here stand for a non-canonical term denoting the ground that corresponds to $[[\mathfrak{R}]]$. We now want to add something to justify the chosen strategy.[38]

First of all, why should one understand a type as a behaviour? A behaviour is a set of designs equal to its bi-orthogonal. This morally means that a behaviour contains all the necessary and sufficient information for a dialogue or game to take place. Thus, a behaviour yields meaningfulness, in the sense that it permits to punctually counter-argue in any possible way over its base.

[37] See Sironi (2014a,b).

[38] Sironi's setup is more complex. Let us sketch it quickly. Given an address ξ, a ramification I and a set of ramifications N, a *positive action* is either $(+, \xi, I)$ or †; a *negative action* is $(-, \xi, N)$. A *chronicle* $\mathfrak{c}$ is a non-empty, finite, alternate sequence of actions such that: (1) each action of $\mathfrak{c}$ is initial or justified by a previous action of opposite polarity; (2) the actions of $\mathfrak{c}$ have distinct addresses; (3) if present, † is the last action of the chronicle. A †-*shorten* of a chronicle $\mathfrak{c}$ is $\mathfrak{c}$ or a prefix of $\mathfrak{c}$ ended by †. The †-*shortening* of E – written $E^\dagger$ – is the set of designs obtained from E by †-shortening chronicles. E is *principal* iff its elements are †-free and $|E^{\perp\perp}| = E^\dagger$. Sironi takes: a type A to be $(P^A)^{\perp\perp}$, with P^A principal; a canonical element of a type to be an element of P^A; a cut-net $\mathfrak{R}$ such that $[[\mathfrak{R}]] \in P^A$ to be a non-canonical element denoting the canonical element that $[[\mathfrak{R}]]$ corresponds to. The appeal to principal set of designs is due to the fact that Sironi aims at having types generated by their canonical elements; indeed, P^A is a kind of minimal generator of $(P^A)^{\perp\perp}$.

From this point of view, a ground is a †-free element of the incarnation of a behaviour for three reasons. It is †-free because, of course, one expects a ground not to involve any paralogism. On the other hand, the incarnation of a behaviour is the subset of the behaviour the elements of which actually operate in interactions. Hence, although we could have demanded a ground to belong to the behaviour as such, we more specifically require a ground to belong to its incarnation so to have something minimal that is used in dialogues or games. Finally, a †-free element $\mathfrak{D}$ of the incarnation $|G|$ of a behaviour G (if any) enjoys the following property: given $\mathfrak{D}' \in |G^{\perp}|$, $[[\mathfrak{D}, \mathfrak{D}']] =$ Daimon, and since $\mathfrak{D}$ is †-free, † must occur in $\mathfrak{D}'$. Hence, we have what we required of designs-as-proofs in Sect. 10.3.3; $\mathfrak{D}$ wins in every possible interaction with elements of the incarnation of the orthogonal of the behaviour which it belongs to. Observe that, if $\mathfrak{D}' \in G^{\perp} - |G^{\perp}|$, $[[\mathfrak{D}, \mathfrak{D}']]$ may diverge; this simply means that $\mathfrak{D}'$ is, so to say, not answering the question raised by $\mathfrak{D}$, i.e. we are not in the presence of an actual dialogue or game.

In the specific case of implication, we thus proceed as follows. We know that a ToG ground over $\mathfrak{B}$ for $\vdash A \to B$ is $\to I\xi^{A}(T)$, where T denotes a constructive function over $\mathfrak{B}$ of type $A \vdash B$. Suppose that we have appropriately translated $\mathfrak{B}$ onto the Ludics framework.[39] Suppose also that the types A and B have been inductively determined as behaviours G^{A} and G^{B}. Observe finally that, as already remarked, interaction, i.e. normalization of cut-nets, is a deterministic procedure.

Let us indicate with $|G^{A}|_{F}$ and $|G^{B}|_{F}$ the sets of the †-free elements of $|G^{A}|$ and $|G^{B}|$ respectively. Given $\mathfrak{D} \in |G^{A}|_{F}$, and given $\mathfrak{D}'$ such that $[[\mathfrak{D}, \mathfrak{D}']] \in |G^{B}|_{F}$, we can take $[[\mathfrak{D}, \mathfrak{D}']]$ as the result of a constructive function of type $A \vdash B$. Hence, we put

$$A \to B = \{\mathfrak{D}' \mid \text{for every } \mathfrak{D} \in |G^{A}|_{F},\ [[\mathfrak{D}, \mathfrak{D}']] \in |G^{B}|_{F}\}^{\perp\perp}.$$

We must take the bi-orthogonal, because we have no guarantee that the above mentioned set is a behaviour. On the other hand, it holds that, for every set of designs E, $E^{\perp}$ is a behaviour. Such technical move has no disturbing effect. We can understand the bi-orthogonal as the interaction-closure of the set as such. Furthermore, the elements of $A \to B$ are designs of base $\alpha \vdash \beta$, where $\vdash \alpha$ is the base of G^{A} and $\vdash \beta$ is the base of G^{B}.[40]

[39] An example of how atomic types can be put in Ludics is given again by Sironi (2014a,b) for the types $\mathbb{N}$ and `List`.

[40] Strictly speaking, what we have defined here is the linear arrow $\multimap$, which is usually done by putting $A \multimap B = A^{\perp} \parr B$. However, our definition is equivalent to the standard one, except that it takes the advantage of not passing through $\parr$ – which would have required the introduction of many new notions and definitions. Observe also that the idea of defining a type as a set closed under bi-orthogonality is quite standard in Linear Logic frameworks, in particular in the Geometry of Interaction program, as remarked in Naibo et al. (2016).

So, $\rightarrow I\xi^A(T)$ can be understood as a †-free element $\mathfrak{D}'$ of $|A \rightarrow B|$. Observe that, for every $\mathfrak{D} \in |G^A|_F$, inductively corresponding to a ground g over $\mathfrak{B}$ for $\vdash A$, the cut-net with $\mathfrak{D}$ and $\mathfrak{D}'$ can be taken as the non-canonical term $\rightarrow E(\rightarrow I\xi^A(T), g)$, denoting a ground over $\mathfrak{B}$ for $\vdash B$. The application of $\mathfrak{D}'$ to $\mathfrak{D}$ produces an interaction $[[\mathfrak{D}, \mathfrak{D}']]$ that ends in an element of $|G^B|_F$.

We quickly outline a concrete – although very simple – example of translation. Consider the orthogonal of the behaviour $\top$ introduced in Sect. 10.3.3, i.e. the behaviour containing only the Daimon. Call it 0. Consider now the following recursive design (called Fax):

$$\dfrac{\dfrac{\dfrac{\vdots\, Fax_{\xi.i \vdash \xi.i}}{\cdots \quad \xi'.i \vdash \xi.i \quad \cdots}\,(\xi', I)}{\cdots \quad \vdash \xi.I, \xi' \quad \cdots}\,(\xi, \mathcal{P}_f(\mathbb{N}))}{\xi \vdash \xi'}$$

Fax is the Ludics interpretation of the identity axiom in sequent calculus. It is a †-free element of the incarnation of any behaviour $A \rightarrow A$. In particular, it is an element of the behaviour $0 \rightarrow 0$. In fact given the only material design in 0 with base ξ – observe that this design *is not* †-free – the normalization between Fax and this latter gives a Daimon based on ξ', which is again a material design in 0. We thus interpret the term $\rightarrow I\xi^0(\xi^0)$ of ToG as Fax in $(0 \rightarrow 0)$. A similar translation may be applied – with some more machinery – to any A, taking a †-free material element of the incarnation of A as inductively corresponding to a ground for $\vdash A$.

However, we remark that, because of the first difference we have highlighted in Sect. 10.4.1 between ToG and Ludics, the indicative mapping we have proposed can work only for *linear* terms of ToG, i.e. terms where every occurrence $\rightarrow I$ binds *exactly one* typed-variable. An extension should take into account what we have already said in the same Section about intuitionistic implication. That is, we need a Ludics interpretation of the modality operator !, and then we can take $A \rightarrow B$ as $(!A) \multimap B$. Since we want designs to be cut-free, this should be done following the work of Faggian and Basaldella[41] rather then Terui with C-Ludics.[42]

If one accepts our reconstruction above of a ToG type as a behaviour, and of a ToG ground as a †-free element of the incarnation of a behaviour, one also easily notices that a type contains more than grounds. Grounds are designs that win in every relevant interaction, but the behaviour they belong to may contain designs with paralogisms or designs losing in some interactions. Thus, we could simply define Cozzo's ground-candidates, discussed in Sect. 10.2.5, as generic elements of a behaviour which are pseudo-ground if they are not †-free or do not belong to the incarnation of the behaviour. In this way, a ground-candidate would be a structure representing one's strategy in a dialogue or game, which might win in some contexts, but lose in others.

[41] See Faggian and Basaldella (2011).

[42] See Terui (2011).

10.5.2 Cozzo's Ground-Candidates Reconsidered

Let us now turn back to Cozzo's fourth objection to the old formulation of ToG discussed in Sect. 10.2.4. Prawitz's definition implied that only valid inferences were inferences, and according to Cozzo this is misleading in that necessity of thought, what Prawitz aims at explaining, can be experienced also in inferences with mistaken premises. To take into account this phenomenon, Cozzo suggests the introduction of ground-candidates, representing a reasoning with possibly mistaken premises, so that a ground-candidate is either an actual ground, or just a pseudo-ground. Cozzo's view seems to involve three aspects that we had better distinguish now:

1. an inference can be performed on wrong premises;
2. an inference performed on wrong premises can be valid;
3. a valid inference on wrong premises can be epistemically compelling.

Point 1 requires that inferential operations are defined, not only on grounds as Prawitz's operations illustrated in Sect. 10.2.2, but also on pseudo-grounds, and hence, more in general, on ground-candidates. Thus, as in point 2, the inference can be valid even if what one is in possession of are just pseudo-grounds. What matters is that, *when applied to grounds* for the premises, the operation always produces a ground for the conclusion.

However, for such an inference to be also compelling, as point 3 requires, it must hold that the inferential operation still produces ground-candidates for the conclusion when applied to pseudo-grounds for the premises. The result of the application might be a ground, but we cannot expect it to be so in every case. And if it were not at least a pseudo-ground, i.e. an epistemic support based on which one feels entitled – perhaps wrongly – to assert the conclusion, there would be no way to maintain that the inference has a compelling character. The question is now whether ground-candidates, in particular pseudo-grounds, can be adequately characterised within the framework of ToG. We shall argue that they cannot.

The problem with ToG is that entities that *are not* grounds cannot reasonably be considered also as pseudo-grounds. This is particularly evident in the case of ground-terms T that denote something which *is not* a ground for $A \vdash B$ – call it $f(\xi^A)$. All we know is that, for some ground g for $\vdash A$, $f(g)$ is not a ground for $\vdash B$. However, for this to be sufficient to conclude that $f(\xi^A)$ is at least a pseudo-ground for $A \vdash B$, it seems plausible to require that $f(g)$ is at least a pseudo-ground for $\vdash B$.

Now, a possibility that may not be excluded is simply that $f(g)$ either diverges, i.e.

$$f(g) = f_1(f(g)) = f_2(f_1(f(g))) = \ldots = f_n(\ldots(f_2(f_1(f(g)))\ldots) = \ldots$$

or gives rise to a loop, i.e.

$$f(g) = f_1(f(g)) = f_2(f_1(f(g))) = \ldots = f_1(f_2(\ldots((\ldots(f_2(f_1(f(g))))\ldots)) = f(g).$$

For example, it has been suggested[43] that this is what happens with paradoxes in a proof-theoretic framework, but without going that far we may consider a "ping-pong" where $f(\xi^A)$ is defined by the equation

$$f(g) = f_1(g, g)$$

and $f_1(\xi^A, \xi^A)$ by the equation

$$f_1(g, g) = f(g).$$

It is clear that we are not entitled to look at $f(g)$ as a pseudo-ground for $\vdash B$. In fact, the computation yields no value, and a computation that yields no value cannot constitute an epistemic support for – possibly mistaken – assertions. As a consequence, a *modus ponens* from $A \to B$ and A to B where one is in possession of $\to I\xi^A(T)$ and g, represented by $\to E(\to I\xi^A(T), g)$, would produce – according to the definition of $\to E$ – the divergent or looping computation $f(g)$. So, in spite of its validity, it would not be compelling. We may postulate, of course, that $f(g)$ is an alleged ground for $\vdash B$, as Prawitz does in response to Cozzo's objection. But, as Usberti remarks, this alleged ground is an entity of any kind, and an assertion based on an entity of any kind can be hardly understood as rational.

Observe that, in the example of divergent or looping computation provided above, we are not only unable to hit upon a ground for $\vdash B$. We are even unable to obtain a *canonical* object of type B. Thus, an obvious way out may be that of requiring that, for it to be a pseudo-ground for $A \vdash B$, $f(\xi^A)$ yields in all cases canonical objects of type B, although for some ground for $\vdash A$ the canonical object produced is not a ground, but only a pseudo-ground for $\vdash B$ – where (some or all) the arguments to which the primitive operation is applied are pseudo-grounds. This should be generalized to ground-candidates, by requiring that a ground-candidate for $A \vdash B$ is a constructive function that, for every ground-candidate for $\vdash A$, yields a canonical object of type $\vdash B$. The notion of ground-candidate can be specified by induction on the complexity of formulas, in the same way as the notion of ground. For example, in the case of implication the clause would run as follows:

($\to^*$) T denotes a ground-candidate for $A \vdash B$ iff $\to I\xi^A(T)$ denotes a ground-candidate for $\vdash A \to B$.

Accordingly, we should modify our notion of inference, by requiring that operations on grounds applied in inference steps are only defined on ground-candidates of the

[43] See Tennant (1982, 1995). For a critical discussion, see, Petrolo and Pistone (2018) and Tranchini (2018).

kind just defined. The *modus ponens* we have considered above, thus, should not be accepted as an inference, since in that case $\rightarrow I$ has an immediate sub-argument that stands for a divergent or looping computation.

In this way, we have excluded from the class of ground-candidates all those entities that are not grounds, and that involve divergent or looping computations. ToG might be modified so as to fit with this idea, but it is doubtful whether this can be done without deeply modifying Prawitz's original project. However, this may constitute the topic of further works. For the moment, we remark that, morally, the result we aim to is exactly what we obtain by understanding Prawitz's grounds as †-free elements of the incarnation of a behaviour and Cozzo's ground-candidates as simple elements of a behaviour. If a design $\mathfrak{D}$ belongs to a behaviour A, it encodes enough information for it to be tested against any design in the behaviour $A^{\perp}$. In particular, this means that $\mathfrak{D}$ has the "shape" of a normal derivation of A, independently of whether such derivation actually exists. The inferential pattern that it is built up is made of primitive steps, independently of whether such steps can be understood as drawn from initial sequents in a specific formal calculus. Finally, it is part of its being the element of a behaviour that any interaction of $\mathfrak{D}$ with any element of $A^{\perp}$ does not give rise to divergent or looping computations. The interaction *produces* a value which may be a ground, but which is not necessarily so. Even in this case, however, the value is still an element of a behaviour, whence it is typed.

10.6 Conclusion

The philosophical and (indicative) formal bridging between Prawitz's theory of grounds and Girard's Ludics we have proposed seems to allow for a dialogical reading of the former. There is still the idea of explaining evidence in terms of some primitive operations that yield normal/canonical objects. Furthermore, evidence is thought of as obtained by performing constructive acts which correspond to reduction/cut-elimination over non-primitive steps. However, evidence now stems from interaction, and since none of the interacting agents might be right, the game may end up in a ground-candidate that happens not to be an actual ground. A picture of this kind seems to be difficult to be accomplished within ToG. As we have seen, attributing types to pseudo-grounds in ToG may be a very difficult task, which requires deep modifications in Prawitz's original project. A type for a pseudo-ground would be nothing but a formal label, and this may not be enough to ensure that the object is constructed by using inferences that are meaning-constitutive, or at least meaning-justifiable. This situation is a priori excluded in Ludics, where we can define a pseudo-ground for A as an element of the behaviour representing A. Elements of the behaviour are abstract counterparts of cut-free derivations for A, and this independently of whether *any* such derivation exists.

Admittedly, our first step in relating ToG and Ludics is very limited. In particular, in order to relate the two theories in a more credible way, one should be able

to regain the full "power" of intuitionistic operations into Ludics. We are quite confident that this can be done in the frame proposed by Faggian and Basaldella[41] Even if Ludics with repetitions does not enjoy all the properties of the version of Ludics we have referred to here, it still does enjoy the ones that are central for our philosophical discussion.

References

Andreoli, J.-M. 1992. Logic programming with focusing proofs in linear logic. *Journal of Logic and Computation* 2(3): 297–347.

Cozzo, C. 1994. *Teoria del significato e filosofia della logica*. CLUEB.

Cozzo, C. 2015. Necessity of thought. In *Dag Prawitz on proofs and meaning*, ed. H. Wansing. Spinger.

Dummett, M. 1993. *The seas of language*. Oxford University Press.

Faggian, C., and M. Basaldella. 2011. Ludics with repetition. exponentials interactive types and completness. *Logical Methods in Computer Science* 7: 1–85.

Fleury, M.R., and M. Quatrin. 2004. First order in ludics. *Mathematical Structures in Computer Science* 14(2): 189–213.

Francez, N. 2015. *Proof theoretical semantics*, volume 57 of *Studies in logic*. College Publication.

Gentzen, G. 1934. Untersuchungen über das logische Schließen I. *Mathematische Zeitschrift* 39: 176–210. Traduction Française de R. Feys et J. Ladrière: Recherches sur la déduction logique, Presses Universitaires de France, Paris, 1955.

Girard, J.-Y. 2001. Locus solum. *Mathematical Structures in Computer Science* 11(3): 301–506.

Howard, W.A. 1980. The formulae-as-types notion of construction. In *To H.B. Curry: Essays on combinatory logic, λ-calculus and formalism*, ed. J. Hindley and J. Seldin, 479–490. Academic Press.

Laurent, O. (2002). *Etude de la polarisation en logique*. Thèse de doctorat, Université Aix-Marseille II.

Martin-Löf, P. 1984. *Intuitionistic type theory*. Napoli: Bibliopolis. (Lecture Notes by G. Sambin).

Martin-Löf, P. 1986. On the meaning of the logical constants and the justification of the logical laws. In *Atti degli Incontri di Logica Matematica vol. 2*.

Naibo, A., M. Petrolo, and T. Seiller. 2016. On the computational meaning of axioms. In *Epistemology, knowledge and the impact of interaction*, volume 38 of *Logic, epistemology, and the unity of science*, 141–184. Springer.

Petrolo, M., and P. Pistone. 2018. On paradoxes in normal form. *Topoi* 38.

Piccolomini d'Aragona, A. 2017. Dag prawitz on proofs, operations and grounding. *Topoi*.

Piccolomini d'Aragona, A. 2018. A partial calculus for dag prawitz's theory of grounds and a decidability issue. In *Philosophy of science. European studies in philosophy of science*, ed. A. Christian, D. Hommen, N. Retzlaff, and G. Schurz, Vol. 9, 223–244. Springer.

Piccolomini d'Aragona, A. 2019. *Dag Prawitz's theory of grounds*. Ph. D. thesis, Aix-Marseille University, "La Sapienza" University of Rome.

Pistone, P. 2018. Polymorphism and the obstinate circularity of second order logic: A victims' tale. *The Bulletin of Symbolic Logic* 24(1): 1–52.

Prawitz, D. 1965. *Natural deduction, a proof-theoretical study*. Number 3 in Acta universitatis stockholmiensis — Stockholm studies in philosophy. Stockholm: Almqvist and Wiksell.

Prawitz, D. 1973. Towards a foundation of a general proof theory. *Studies in Logic and the Foundations of Mathematics* 74.

Prawitz, D. 1977. Meaning and proofs: On the conflict between classical and intuitionistic logic. *Theoria* 43: 2–40.

Prawitz, D. 2009. Inference and knowledge. In *The logica yearbook 2008*, ed. M. Pelis. College Publications.
Prawitz, D. 2012. The epistemic significance of valid inference. *Synthese* 187(3): 887–898.
Prawitz, D. 2015. Explaining deductive inference. In *Dag Prawitz on proofs and meaning*, ed. H. Wansing, 65–100. Spinger.
Schroeder-Heister, P. 2006. Validity concepts in proof-theoretic semantics. *Synthese*.
Schroeder-Heister, P. 2009. Sequent calculi and bidirectional natural deduction: On the proper basis of proof-theoretic semantics. In *The logica yearbook 2008*, ed. M. Pelis. College Publications.
Sironi, E. 2014a. Type theory in ludics.
Sironi, E. 2014b. *Types in Ludics*. Ph. D. thesis, Aix-Marseille University.
Sundholm, G. 1998. Proofs as acts and proofs as objects. *Theoria*.
Tennant, N. 1982. Proof and paradox. *Dialectica*.
Tennant, N. 1995. On paradox without self-reference. *Analysis* 55(3): 199–207.
Terui, K. 2011. Computational ludics. *Theoretical Computer Science* 412(20): 2048–2071.
Tranchini, L. 2014a. Dag prawitz. *APhEX* 9.
Tranchini, L. 2014b. Proof-theoretic semantics, paradoxes and the distinction between sense and denotation. *Journal of Logic and Computation* 26.
Tranchini, L. 2018. Proof, meaning and paradox: Some remarks. *Topoi* 38: 1–13.
Troelsta A., and D. Van-Dalen. 1988. *Constructivism in mathematics vol. I*. Nort-Holland Publishing Company.
Usberti, G. 2015. A notion of c-justification for empirical statements. In *Dag Prawitz on proofs and meaning*, ed. H. Wansing, 415–450. Spinger.
Usberti, G. 2017. Inference and epistemic transparency. *Topoi* 1–14.
Zeilberger, N. 2008. On the unity of duality. *Annals of Pure and Applied Logic* 153: 66–96.

Chapter 11
Predicativity and Constructive Mathematics

Laura Crosilla

Abstract In this article I present a disagreement between classical and constructive approaches to predicativity regarding the predicative status of so-called generalised inductive definitions. I begin by offering some motivation for an enquiry in the predicative foundations of constructive mathematics, by looking at contemporary work at the intersection between mathematics and computer science. I then review the background notions and spell out the above-mentioned disagreement between classical and constructive approaches to predicativity. Finally, I look at possible ways of defending the constructive predicativity of inductive definitions.

Keywords Inductive definitions · Predicativity · Constructive mathematics · Vicious circle principle · Invariance

11.1 Introduction

Constructive mathematics is a form of mathematics which uses *intuitionistic* rather than classical logic. Different varieties of mathematics based on intuitionistic logic have been proposed over the years since Brouwer's inception of intuitionism. In the following, "constructive mathematics" denotes "Bishop–style" mathematics, the mathematics based on intuitionistic logic initiated by Errett Bishop in "Foundations of constructive analysis" (Bishop, 1967).[1] Constructive mathematics has since witnessed substantial advances in analysis, topology and algebra. Starting from the 1970s, a number of formal systems have been proposed to codify or formalise this form of mathematics. Their aim was to isolate the principles underlying constructive

[1] See Bridges and Richman (1987) for an introduction.

L. Crosilla (✉)
Department of Philosophy, IFIKK, University of Oslo, Blindern, Norway
e-mail: Laura.Crosilla@ifikk.uio.no

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_11

mathematics' fundamental concepts, especially its concepts of set and function. Among these systems are Martin-Löf Type Theory and Constructive Set Theory.[2]

In this article, I consider one aspect in which constructive and classical foundations of mathematics differ. Constructive set and type theories diverge from standard classical set theories such as Zermelo-Fraenkel set theory in two distinct respects: they employ *intuitionistic* rather than classical logic and they comply with a form of *predicativity*. Predicativity is my main objective, as I compare a prominent classical approach to predicativity with the form of predicativity that we find in constructive foundational systems. My focus is therefore not a comparison between constructive systems and standard classical foundations such as ZFC, rather a comparison between two distinct proposals for developing mathematics on the basis of a predicative concept of set. In so doing, I shed some light on the very notion of predicativity constructive systems contrive, which is not fully spelled out in the relevant literature.

In the following, I focus on a disagreement between standard classical and constructive approaches to predicativity. This regards the predicative status of so-called *generalised* inductive definitions. An inductive definition defines a set, say X, by: (i) identifying some *initial elements* of X; (ii) specifying new elements of X *in terms of elements already included in it* and (iii) finally adding that *nothing else* belongs to X. Inductive definitions are clearly appealing from a constructive point of view, as they present a set as if it were constructed step-by-step from below. The metaphor of a step-wise generation of a set from below is also often employed to convey the notion of *predicative definition* of a set, i.e. of a definition that is not viciously circular. While salient inductive definitions are considered constructive, their predicative status is disputed. According to constructive approaches to predicativity, such as the one developed in Martin-Löf type theory, generalised inductive definitions are acceptable. They are, however, impredicative according to a well-known classical approach to predicativity.[3]

The remarkable feature of this disagreement is that constructive approaches to predicativity may be seen as more "generous" compared with standard classical approaches to predicativity. This fact is at first sight surprising, since we usually expect constructive foundations to be more restrictive than their classical counterparts. Constructive foundations are indeed substantially more restrictive than impredicative foundations such as ZFC, in the sense that they do not countenance

[2] See e.g. Martin-Löf (1975), Myhill (1975), Aczel (1978), Martin-Löf (1984), Beeson (1985). Another approach to the foundations of constructive mathematics is Feferman's Explicit Mathematics, which has been studied especially in proof theory (Feferman, 1975). A very recent development is Homotopy Type Theory (Univalent Foundations Program, 2013).

[3] As further clarified in Sect. 11.4.1, the debate on the predicative status of inductive definitions has focused on *generalised* inductive definitions. In the following, unless otherwise stated, I omit the qualification "generalised".

impredicative and essentially non-constructive methods of proof.[4] However, if we consider *predicative* approaches to foundations, the standard classical approach elaborated, for example, by Kreisel, Feferman and Schütte turns out to be more restrictive compared with the constructive one, and the key difference is its rejection of generalised inductive definitions. This observation can be made precise by employing fundamental results in ordinal analysis, a branch of proof theory. By carefully assigning ordinals to formal theories, proof theorists have devised means of comparing theories in terms of their "proof-theoretic strength". An outcome of that research is that constructive theories such as Martin-Löf Type Theory and Constructive Zermelo Fraenkel set theory countenance systems which are proof-theoretically much stronger than classical theories that have been devised to codify (classical) predicativity. Crucially, these constructive systems have the resources to express generalised inductive definitions.[5]

This disagreement between classical and constructive forms of predicativity is pivotal for an understanding of predicativity and for an assessment of its significance within the foundations of mathematics. Inductive definitions play a substantial role within the contemporary constructive practice, as they are a fundamental component of constructive sets and type theories and also play a major role within constructive proof assistants such as Coq (The Coq Development Team, 2020). For this reason, an analysis of the foundational status of these definitions is timely and valuable. Surprisingly, the relevant literature does not offer, as far as I know, a sharp delineation of the notion of constructive predicativity. More specifically, there is no detailed philosophical comparison between classical and constructive forms of predicativity nor an analysis of the above-mentioned disagreement between classical and constructive forms of predicativity. As (generalised) inductive definitions are considered a crucial component of constructive predicativity, an analysis of this disagreement between classical and constructive forms of predicativity is bound to shed light on the very notion of constructive predicativity and contribute to a more precise characterisation of this notion. For these reasons, a fundamental step into a philosophical investigation of constructive predicativity has to be an explication of the disagreement between classical and constructive approaches to predicativity over the status of generalised inductive definitions.

In the first part of this article, I begin by offering some motivation for an inquiry into the predicative foundations of constructive mathematics, by looking at contemporary work at the intersection between mathematics and computer science. I then review the background notions and spell out the above-mentioned disagreement between classical and constructive approaches to predicativity. In the second part of this article, I look at possible ways of defending the constructive predicativity of

[4] Arguably, from a different perspective, constructive systems are more flexible and less restrictive than traditional classical systems such as ZFC, as they allow for a variety of interpretations, including computational interpretations (see Sect. 11.2). See also Bridges and Richman (1987).

[5] See Martin-Löf (1984), Aczel (1986), Palmgren (1992), Dybjer (2000), and Dybjer and Setzer (2003).

inductive definitions. Due to space constraints, I can only quickly sketch the main ideas. My proposal is to explore and further expand ideas on predicativity first put forth by Poincaré and Weyl at the turn of the twentieth century, as they seem to offer a plausible route leading to the claim that inductive definitions are predicative from a constructive perspective. I also highlight the importance of clarifying whether the underlying logic, classical or intuitionistic, may have a role to play in assessing the predicative status of inductive definitions. A full assessment of the complex question of whether generalised inductive definitions are constructively predicatively justified but classically predicatively inadmissible will have to be postponed to another occasion. My hope at present is to generate some discussion on this important question and, more generally, on the very notion of constructive predicativity.

11.2 Motivation: Constructive Mathematics as Algorithmic Mathematics

One of constructive mathematics' most significant characteristics is that its theorems afford a computational or "algorithmic" interpretation: they can, at least in principle, run on a computer. Bishop's pioneer realisation that the exclusive use of intuitionistic logic could endow mathematical theorems with computational meaning has been vindicated in recent years. In fact, constructive mathematics and, especially, constructive type theory, have been fundamental source of inspiration for the theory and the applications of computer aided mathematics. One of the main instruments in this thriving area of research are *proof assistants*, i.e. computer software which is used interactively to formalize mathematical proofs. In recent times large and complex proofs of mathematical theorems, such as the Four Colour Theorem in graph theory and the Feit-Thompson Theorem in finite group theory have been implemented in such systems.[6]

Proof assistants are primarily used to completely formalize proofs and check their correctness. This is no trivial work, as a thorough formalization of a straightforward theorem requires not only to fill in all the gaps routinely left out in an informal proof and correct possible mistakes, but also formalize substantial portions of mathematics in view of all the background definitions and results the theorem depends on. It also involves subtle choices on how to best formalize individual components of a proof. In addition to this "primary" application within mathematics the formalization of mathematical proofs has other uses, which are attracting renewed interest for this area of research. For example, proof assistants are also applied to verify the correctness of computer software. A further emerging area of research looks at utilizing proof assistants to "extract" computer programs from fully formalized

[6] See, for example, Martin-Löf (1982), Coquand and Huet (1986), Constable and et al. (1986), Nordström et al. (1990), Gonthier (2008), AGDA (2020), and The Coq Development Team (2020).

proofs. Here constructive proofs have been the main focus so far, as we can make use of their interpretation as algorithms to produce real-life, working programs.

The extensive research on proof assistants, motivated as it is by a number of applications, is bound to have a considerable impact within mathematics. Mathematical proofs are becoming increasingly complex and large. Computer systems that check the correctness of proofs are therefore likely to become a significant part of everyday mathematics. The hope is that computer systems could over time help us not only check existing large proofs, but also find new ones and develop effective proof strategies. Since constructive type theories (both predicative and impredicative) are at the heart of some of the most widely used proof assistants (e.g. Coq) these new developments may change significantly the perceived position of constructive mathematics within the mathematical community, granting it a more central role. For these reasons it is necessary that the philosopher of mathematics reflects on constructive mathematics and its philosophical motivations and compares it with the better-known classical practice.

Predicativity is a crucial component of foundational systems such as Martin-Löf type theory. The form of predicativity that we find in this theory combines the availability of a quite general form of inductive definitions (e.g. in the guise of so-called W types) with a strong form of Curry-Howard correspondence.[7] The latter endows the logical constants with a direct computational meaning which is key to the theory's role as programming framework, as clarified in Martin-Löf (1982). This interpretation of the logical constants, however, turns out to be incompatible with impredicativity, as demonstrated by Girard's paradox.[8] Martin-Löf's way out of paradox was to abide to a form of predicativity while enriching the theory with powerful type constructing devices, i.e. W types and reflecting universes.[9]

Predicativity is widely discussed from a *technical* point of view also within the Coq community. While the calculus of constructions (i.e. the type theory on which the Coq system is founded) features a strong form of impredicativity, recent versions of Coq have restricted this impredicativity so to gain more flexibility and ease compatibility with mainstream mathematics. Given the varieties of applications of proof assistants, it is important to allow for the possibility of adding assumptions that enable the formalization of different forms of mathematics, such as, for example, the axiom of choice or the principle of excluded middle, which are required to formalise standard classical mathematics. It is here that the notion

[7] The W type constructor is used to codify well-founded trees in type theory. It can therefore be used to codify Brouwer's constructive ordinals (see Sect. 11.4.1). The Curry-Howard correspondence, also known as "formulas-as-types" correlates intuitionistic logic with type theories. See e.g. Troelstra (1999) for details and Crosilla (2019) for an informal discussion of its relation with predicativity.

[8] This paradox affected an early variant of type theory, which included a type of all types. See Girard (1972). See also Coquand (1989) and Martin-Löf (2008) for analysis and Crosilla (2019) for philosophical reflections.

[9] Universes in Martin-Löf type theory are powerful constructs which act as reflection principles. Roughly a universe is a type closed under certain type-forming operations.

of predicativity (in the form of syntactic constraints that block specific forms of impredicativity) has proved useful.[10]

As (generalised) inductive definitions are increasingly employed within the computer aided formalization of mathematics and are considered predicative there and within constructive mathematics, a clarification of their predicative status becomes particularly urgent. In this context, the disagreement between alternative approaches to predicativity that was mentioned in the Introduction becomes particularly significant. In the next section, I review the standard classical approach to predicativity which emerged from fundamental work in proof theory, before turning to the constructive case in subsequent sections.

11.3 Predicativity Given the Natural Numbers: The Classical Approach

The notion of predicativity emerged at the beginning of the last century within Poincaré and Russell's analysis of the set-theoretic paradoxes.[11] The analysis identified a form of vicious circularity as source of the paradoxes. This circularity is manifested in problematic *impredicative* definitions which attempt to define mathematical entities in a circular way, e.g. by specifying an element of a collection in terms of *all* the elements of that collection. Adherence to predicativity was therefore proposed as an instrument for avoiding vicious circularity in definitions and, in this way, stay clear of paradoxes. Russell introduced his well-known *"Vicious–Circle Principle" (VCP)*, according to which no totality can contain members only definable in terms of this totality.[12] Russell's technical solution to the difficulty was *ramified* type theory, designed to ensure full compliance with the VCP.[13] The main idea of ramified type theory is to define sets (i.e. types) by introducing simultaneously two kinds of regimentation: type levels and orders. The latter regiment propositional functions so to ensure that properties defined in terms of the totality of properties of a given order belong to the next higher order. The interplay of these restrictions aims at avoiding the occurrence of the perceived problematic circularity in definitions. Avoiding vicious circularity in analysis was

[10] In Coq there are two sorts (i.e. categories) of objects "Prop" and "Set". Both had impredicative features in early versions of the system, so that, for example, one could quantify over all Sets to define a new set. Recent versions, however, retain an impredicative "Prop" but abandon the impredicativity of "Set". These new restrictions are introduced to increase compatibility with classical mathematics (see e.g. Barbanera and Berardi (1996)).

[11] See, for example, Poincaré (1905, 1906a,b), Russell (1906a,b, 1908) and Poincaré (1909, 1912).

[12] Russell and Whitehead gave a number of renderings of the VCP. For example, *"no totality can contain members defined in terms of itself"* (Russell, 1908, p. 237) and "[. . .] whatever in any way concerns *all* or *any* or *some* of a class must not be itself one of the members of a class" (Russell, 1973, p. 198). See also Gödel (1944) for an influential discussion, especially p. 454-5.

[13] See Russell (1908) and Whitehead and Russell (1910–1913).

also Weyl's aim in "Das Kontinuum" (Weyl, 1918), where a highly original and influential predicative treatment of analysis was undertaken without recourse to ramification.[14]

Following Poincaré and Russell, (im)predicativity is usually characterised as follows:

(i) a definition is impredicative if it defines an entity in terms of a totality to which the entity itself belongs; it is predicative otherwise.
(ii) a mathematical entity (e.g. a set) is impredicative if it can *only* be defined by an impredicative definition; it is predicative otherwise.

The qualification "only" in clause (ii) is important. This clause states that an entity is predicative provided that it affords a predicative definition. Since it is common for a mathematical entity to be defined in a number of equivalent ways, an entity is considered impredicative as long as no alternative predicative definition of it is available. Extensive work in mathematical logic in recent years has shown that many apparently impredicative notions in analysis can be reformulated so to afford predicative treatment.[15] This work is complex, as one typically needs to re-frame one's definitions to avoid impredicativity. Sometimes this requires the redevelopment of a substantial portion of mathematics. Weyl's book "Das Kontinuum" (Weyl, 1918) sets out a fundamental example in this respect, as it shows how to carry out large portions of analysis from a predicative point of view.

Of special interest in the present context are developments that took place from the 1950s, when prominent logicians undertook a precise formal analysis of predicativity. Only the most general points of that development are needed for the present discussion.[16] Among these new developments, one course of thought brought to what is often termed "predicativity given the natural numbers" (Feferman, 2005). From a technical point of view, this may be seen as a continuation of both Russell's ramified type theory and Weyl's predicative analysis. It takes the VCP as the main guiding principle and further develops Russell's idea of ramification. It also takes the natural numbers as 'given', while introducing predicatively motivated constraints on subsets of the natural numbers, as Weyl did. The thought is that the natural numbers are unproblematic and safe, but sets of natural numbers need to be defined predicatively to avoid vicious circularity. Importantly, as in Weyl's (1918) booklet and in Russell's type theory, one uses *classical logic* throughout.

Notwithstanding these similarities, there is a significant difference with Weyl's predicativism. The aim of the logical analysis of predicativity was not a predicative foundation of analysis, with the consequent abandonment of those parts of analysis that could not be rephrased in purely predicative terms. The main focus was rather a clarification –from the outside so to speak– of the limit of predicativity: how far

[14] See Sect. 11.4.3.1 for more on Weyl (1918).

[15] See e.g. Feferman (1988, 2004, 2013b) and Simpson (1988, 1999).

[16] Some of the most significant steps in that development are recalled in Feferman (2005). See also Dean and Walsh (2016) and Crosilla (2017).

can we reach if we take a predicativist stance? This question was approached along two main dimensions: (1) by using mathematical logic to determine the limit of predicativity and (2) by a case by case study of ordinary mathematics to assess which parts of it can be given predicative treatment. A further difference with Weyl is that the new attitude as well as the more refined logical instruments in the meantime available brought the logicians to go beyond Weyl's predicative analysis, by contemplating transfinite iterations of ramified comprehension along so-called predicative ordinals. Through fundamental contributions by Kreisel, Feferman and Schütte the "logical analysis of predicativity" gave rise to an exemplary chapter in proof theory, which culminated with the determination by Feferman and Schütte (independently) of the *limit* of predicativity by means of ordinal analysis.[17]

Ordinal analysis uses proof-theoretic techniques to assign ordinals to theories as a way of assessing and comparing their strength. The proof-theoretic analysis of predicativity of the 1960s made use of a transfinite hierarchy of subsystems of second order arithmetic with ramified comprehension (also called ramified analysis). The main idea is that each system allows for a ramified form of comprehension, thus only "referring" to entities populating earlier stages of the hierarchy. This ensures that each level of the hierarchy is predicatively justified. Crucially, the hierarchy is indexed by ordinals and a substantial contribution of this analysis was a proposal on how far along the ordinals we may proceed without stepping into impredicativity. To this end, the notion of predicatively provable ordinal was introduced with the intention to capture the concept of an ordinal a predicativist would *recognize*. Roughly, predicatively provable ordinals can be defined "from below" through a bootstrapping process: one progresses along the ramified hierarchy to a theory indexed by an ordinal α only if one has already proved that α is an ordinal in a "previous" theory within the hierarchy. This hierarchy of formal systems then acted as canonical reference: one considers predicative any formal system which can be reduced to a system in that hierarchy (according to a formally specified notion of proof-theoretic reduction). The so-called limit of predicativity was then identified in terms of an ordinal known as Γ_0, the first non-predicatively provable ordinal.[18]

[17] See Kreisel (1958), Feferman (1964), and Schütte (1965a,b). Note that this is not the only logical analysis of predicativity proposed in the 1950–1960s. Another approach (Kreisel, 1960) made essential use of work in recursion theory and definability theory, and identified the predicatively definable sets of natural numbers with the so-called hyperarithmetical sets. Here work by Kleene, among others, provided fundamental insights and the necessary tools for the analysis. See Moschovakis (1974) for the relevant notions, historical notes and references.

[18] Schütte's fundamental contribution to this analysis of predicativity is acknowledged by Feferman (2013a, p. 8–9) as follows: "[...] the determination by Schütte and me in the mid-1960s of Γ_0 as the upper bound for the ordinal of predicativity simply fell out of his ordinal analysis of the systems of ramified analysis translated into infinitary rules of inference when one added the condition of autonomy."

The logical analysis of predicativity therefore made a clear and precise proposal for a formal analysis of predicativity, employing state of the art logical machinery to extend Russell's and Weyl's work.

11.4 Constructive Predicativity and Inductive Definitions

A number of formal systems have been introduced over the years to formalise constructive mathematics. It is common to distinguish two kinds of systems: impredicative systems such as Intuitionistic Zermelo Fraenkel set theory and the Calculus of Constructions (Friedman, 1973; Coquand and Huet, 1986), and predicative systems, such as Martin-Löf Type Theory and Constructive Zermelo Fraenkel set theory (Aczel, 1978; Martin-Löf, 1975). While the latter theories are said to be predicative, the literature does not offer a sharp delineation of the relevant notion of predicativity, nor is there an authoritative analysis of this notion comparable to the insightful appraisal given over the years in the classical case, especially through the work of Feferman.[19] There is agreement among constructive mathematicians on paradigmatic examples of impredicativity: as in the classical approach to predicativity discussed above, the powerset of an infinite set is considered impredicative, and so is full second order arithmetic. Other forms of higher order quantification (e.g. over so-called "propositions" in type theory) are also considered impredicative. Moreover, there is overall agreement in the literature that some generalised inductive definitions are constructively predicatively justified.[20] In fact, the acceptance of (at least some) generalised inductive definitions is often taken to be the main characteristic distinguishing the constructive from the classical approach to predicativity discussed in the previous Section. In fact, according to classical predicativity given the natural numbers generalised inductive definitions are impredicative, on the basis of proof-theoretic results.

In the following, I first review the notion of generalised inductive definition and then investigate why it is considered problematic from a classical predicativist perspective but may be considered unproblematic from a constructive point of view.

[19] For discussion see Coquand (1989), Dybjer (2012), Palmgren (1998) and Rathjen (2005).

[20] Note that while my focus in this note are intuitionistic theories, Lorenzen and Myhill have argued for a rather liberal notion of predicativity with respect to a quite general notion of constructivity (also in the context of theories with classical logic). See especially (Lorenzen, 1958; Lorenzen and Myhill, 1959). See also Wang (1959). For Martin-Löf type theory, see e.g. Palmgren (1998) and Rathjen (2005).

11.4.1 Inductive Definitions

Inductive definitions were used in constructive mathematics from the very start as witnessed, for example, by Brouwer's constructive ordinals. In mathematical logic, inductive definitions gained particular relevance from the 1950s especially in recursion theory and in proof theory.[21] A principal reason for the focus on inductive definitions in proof theory was Kreisel's hope that the study of formal theories for inductive definitions could clarify whether Spector's 1961 proof of consistency of second order arithmetic could be constructively justified. In fact, it turned out that such theories are not sufficiently strong to accomplish this task, as their proof-theoretic strength is strictly in between the strength of predicative theories (according to the Γ_0 analysis) and full second order arithmetic. However, the proof-theoretic investigation of theories of inductive definitions gave rise to crucial advances in ordinal analysis and was also key to the proof-theoretic study of prominent impredicative subsystems of second order arithmetic.[22]

Today inductive definitions figure prominently in Martin-Löf type theory, for example in the form of well-founded trees, which are defined by employing the so-called **W** type constructor. In the case of type theory, the combination of well-founded trees and universes (i.e. reflection principles) endow this theory with considerable proof-theoretic strength, well exceeding the strength of theories in the ramified hierarchy up to Γ_0. Martin-Löf type theory therefore includes systems whose proof-theoretic strength well exceeds the realm of predicativity given the natural numbers.[23] Recent years have also seen frequent application of inductive definitions in constructive mathematics. For example, they have been successfully employed in formal topology (Coquand et al., 2003; Sambin, 1987) to circumvent the ubiquitous use of the powerset operation. As already mentioned in Sect. 11.2, inductive definitions are also extensively used in the formalization of mathematics within theorem provers such as Coq. One reason for this is that inductive definitions offer a uniform way of characterising a number of type constructions, avoiding the proliferation of primitive types. For example, given a general scheme for inductive definitions, one can apply it to define the natural numbers, without assuming a primitive type of natural numbers.[24]

An inductive definition defines a set, say X, by identifying some initial elements of it and specifying all the remaining elements of X in terms of elements already included in it. It may be helpful to see how one usually characterises inductive definitions from a standard set-theoretic perspective. Here an inductively defined

[21] See the fundamental (Barwise, 1975; Moschovakis, 1974).

[22] See Buchholz et al. (1981). The introduction gives an insight into the historical developments of ordinal analysis beyond predicativity. See Chapter 1 for background. See also Feferman (2013a); Martin-Löf (2008).

[23] See e.g. Palmgren (1992, 1998), Rathjen et al. (1998), Dybjer (2000), Dybjer and Setzer (2003) and Rathjen (2005).

[24] See Dybjer (2012) for discussion and references.

set may be seen as the least fixed point of a monotone operator. For our purposes, it suffices to focus on the case of inductive definitions of sets of natural numbers.[25] Let $\Gamma : P(\mathbb{N}) \to P(\mathbb{N})$ be an operator (or function) from the power set of the natural numbers to the power set of the natural numbers and let $X, Y \in P(\mathbb{N})$. We say that Γ is **monotone** if:

$$X \subseteq Y \to \Gamma(X) \subseteq \Gamma(Y).$$

X is Γ**-closed** when

$$\Gamma(X) \subseteq X.$$

For monotone Γ, it is easy to show that there is the smallest Γ-closed set, also called the *fixed point* of Γ:

$$I_\Gamma = \bigcap \{X : X \text{ is } \Gamma\text{-closed}\}.$$

The constructive appeal of inductive definitions is due to the fact that they can be thought of as constructing a set step-by-step and from below. This metaphor of a bottom-up construction can be made more precise by using the ordinals to index the stages of the least fixed point of a monotone operator. One starts from stage 0 and successively applies the operator Γ to go from one stage to the next. More precisely, given Γ as above, and I_Γ its least fixed point, the α-stage of I_Γ is:

$$I_\Gamma^\alpha = \Gamma(\bigcup_{\beta<\alpha} I_\Gamma^\beta).$$

The crucial point is that, at the price of taking the classical ordinals as given, an inductively defined set can now be presented as the closure of a step-by-step process of generation, so that each stage is the result of applying the operator to a previously generated fragment of the set.

The reference above to the classical ordinals, however, is problematic from a constructive perspective. Another way of presenting inductive definitions may, however, be more appealing from a constructive perspective. This is in terms of a set of rules that specify the elements of an inductively defined set. Typically, one starts from some initial elements and then gives rules that yield new elements of a set from "previously constructed" elements of it. The least set closed under these rules is then the set inductively defined by them.[26]

[25] See the exposition in Buchholz et al. (1981), Chapter 1.

[26] See Aczel (1977). Particularly appealing from a constructive point of view are deterministic rules. A rule is deterministic if for any conclusion a there exists exactly one set of premises X such that a is a consequence of X according to the rule.

The simplest example of inductive definition of an infinite set is the inductive definition of the set of natural numbers as the smallest set containing 0 and closed under the successor operation. One has the following introduction rules:

1. 0 is a natural number,
2. if n is a natural number, then its successor, $suc(n)$, is also a natural number.

Taking the natural numbers to be the *smallest* set satisfying these rules, amounts to adding the claim that *nothing else* is a natural number. The latter clause is expressed by the principle of mathematical induction, which is often formulated as an elimination rule complementing the introduction rules.

The example of the inductive definition of the natural numbers is here chosen for its simplicity. As already mentioned, both predicativity given the natural numbers and constructive predicativity take the natural numbers as unproblematic, as given, and introduce predicatively motivated constraints on subsets of the natural numbers.[27] The inductive definitions that rise concerns from a standard (classical) predicativist perspective are those that go beyond the natural numbers, like, for example, the definition of the constructive ordinals. The latter can be defined by the following introduction rules:

1. 0 is in $\mathcal{O}$,
2. if a is in $\mathcal{O}$, then $suc(a)$ is in $\mathcal{O}$,
3. if f is a function from the natural numbers, $\mathbb{N}$, to $\mathcal{O}$ and for all n in $\mathbb{N}$, $f(n)$ is in $\mathcal{O}$, then the supremum of the $f(n)$ is in $\mathcal{O}$.

While the inductive definition of the natural numbers is finitary, in the sense that each rule has only finitely many premises, the definition of the constructive ordinals is an example of infinitary inductive definition. Note also that the definition of the constructive ordinals builds on the definition of the natural numbers. One can further iterate this process and build a new inductive definition on the basis of the constructive ordinals, and so on. In this way, the so-called "higher tree classes" can be defined inductively.[28] *Generalised* inductive definitions include definitions such as that of $\mathcal{O}$ and also countenance iterated inductive definitions.

The proof theory of inductive definitions has focused on formal theories that codify generalised inductive definitions. These formal theories extend Peano Arithmetic by introducing predicates for so-called positively definable operators. Here the positivity of the relevant predicates is required to ensure the monotonicity of the operators they define.[29] Theories, based on intuitionistic logic (i.e. extensions

[27] Note that while the forms of predicativity considered in this article take the natural numbers as unproblematic, this assumption is not gone unchallenged. Dummett, Nelson and Parsons have (independently) argued for the impredicativity of the principle of mathematical induction (Dummett, 1963; Nelson, 1986; Parsons, 1992). Nelson (1986) develops a form of predicative arithmetic that substantially constrains mathematical induction, therefore giving rise to weak subsystems of Peano Arithmetic.

[28] See Buchholz et al. (1981, p. 147).

[29] See Buchholz et al. (1981, Chapter 1).

of Heyting Arithmetic) have also been considered and have played a crucial role in the proof theoretic analysis. In practice, theories of inductive definitions have acted as systems of reference in the proof-theoretic analysis of inductive definitions, therefore playing a similar role in this context as systems of ramified analysis for the proof-theoretic analysis of predicativity. A well-known theory which formalises non-iterated inductive definitions goes under the name of ID_1. Stronger theories have been introduced to codify iterated inductive definitions. As already mentioned at the beginning of this section, the proof-theoretic analysis of theories of inductive definitions shows that their proof-theoretic strength exceeds that of predicative theories according to the Γ_0 analysis (Buchholz et al., 1981). This is the case already for the theory ID_1, whose proof-theoretic ordinal, the so-called Bachmann-Howard ordinal, is much larger than Γ_0.

11.4.2 The Impredicativity of Generalised Inductive Definitions

Generalised inductive definitions are considered impredicative according to the logical analysis of predicativity mentioned in Sect. 11.3. The worry is that in the build up of an inductive set we need to refer to the very set we are defining, thus contravening the VCP. In the terminology introduced in the previous section, the main difficulty lies in the claim that the inductively defined set we are defining is the *least* fixed point of the given inductive definition. This worry is particularly evident when we look at the standard set-theoretic presentation of inductive definitions. If we take the set-theoretic definition of the *least* fixed point of an inductive definition as the intersection of all Γ-closed subsets of $\mathbb{N}$, for some monotone operator Γ, then the difficulty is obvious: we define a subset of the natural numbers by reference to a collection of subsets of the natural numbers to which it belongs, against the VCP.

Arguably, one of the main benefits of the logical analysis of predicativity is that it has revealed that apparently impredicative notions of ordinary mathematics could after all be given a predicative treatment. As a consequence, a prima facie impredicativity could be eliminated. We could then explain a prima facie impredicativity as a by-product of its codification within a certain conceptual framework (e.g. set theory). One could hope that similar considerations could also be applied to the case of inductive definitions: while the set-theoretic framework strongly suggests the impredicativity of inductive definitions, a more careful analysis could perhaps offer a different verdict (at least in the case of the inductive definitions the constructivist cares about). For example, one could hope that an idea mentioned towards the end of the previous section could help defuse the impredicativity of inductive definitions. There we saw that the classical ordinals can be used to index the stages of the least fixed point of an inductive definition. The ordinals, in other terms, can help us stratify an inductively defined set so that at each step we refer only to "previously" constructed fragments of it. Borrowing the proof-theorist's terminology (Buchholz et al., 1981, p. 262-3), with the help of the ordinals, an inductive definition such as that of $\mathcal{O}$ can be expressed in such a way that it becomes *locally* predicative, i.e.

it is predicative at each stage since, locally, we only refer to what has already been constructed, rather than to the whole set under construction. However, the difficulty with this strategy is that the ordinals we need to employ in order to index the stages of the inductive definition cannot be given a predicative justification that would satisfy the predicativist given the natural numbers. This is confirmed by the proof-theoretic analysis of theories of inductive definitions which are seen to exceed the proof-theoretic strength of predicative theories. In other terms, one could claim that we have local predicativity, but impredicativity as a whole.

Perhaps representing inductive definitions in terms of rules, with no explicit mention of the classical ordinals, could help explain away their impredicativity. The worry in this case is that the rules themselves may involve a circularity. In the case of generalised inductive definitions, in fact, the clause expressing the minimality of the inductive definition will have no restriction to prevent it from referring to the very set it inductively defines.[30]

These observations can be made fully precise through a careful proof-theoretic analysis of formal theories for inductive definitions. The main "argument" adduced for the impredicativity of inductive definitions, therefore, is the fact that the proof-theoretic strength of theories of inductive definitions exceeds the limit of predicativity given the natural numbers, captured by the ordinal Γ_0. In fact, there is no proof-theoretic reduction of theories of (generalised) inductive definitions to the systems of ramified analysis, since the former are proof-theoretically much stronger than the latter. As we saw above, the ramified hierarchy acts as canonical systems of reference for predicativity given the natural numbers. Therefore, the fact that the proof-theoretic strength of theories of inductive definitions exceeds the strength of the whole ramified hierarchy is taken as clear indication that generalised inductive definitions involve impredicativity.

11.4.3 Predicative After All?

Although inductive definitions are considered impredicative according to predicativity given the natural numbers, they are usually considered constructive and predicative in the constructive literature. The term "constructive" is notoriously vague and is routinely applied to a variety of forms of mathematics, often very different from each other. It is thus perhaps not that surprising that the literature presents us also with the claim that generalised inductive definitions are constructive, but impredicative. The availability of a set of *rules* for the generation of the elements of an inductively defined set is often considered key to the *constructivity* of inductive definitions. For example, when inductively defining an infinite set,

[30] A "miniature" argument along these lines can be carried out already in the case of the natural numbers to argue for the impredicativity of the induction principle. This will be discussed in Sect. 11.4.3.2.

one does so by means of fixed rules and in a uniform way: by employing some initial elements of the set and repeatedly applying a uniform procedure to obtain all the other elements of the set. Crucially, the induction principle associated with an inductive definition somehow mirrors the construction of the elements of the set. Hence, the proofs are also likewise structured. This latter point is rightly stressed by Sieg (Buchholz et al., 1981, Chapter 3, p. 147), when discussing the intuitionistic theory that formalises the construction over Heyting Arithmetic of the constructive ordinals. Sieg writes that this theory is constructively justified:

> and by that I simply mean that the theory is based on intuitionistic logic, the objects in its intended model are exhibited or obtained by construction and the proof-procedures follow or parallel the construction of the objects.

While it is usually agreed that (at least some) generalised inductive definitions are constructive, it is their predicativity that is controversial (Buchholz et al., 1981; Feferman, 1964). For example, the inductive definition of $\mathcal{O}$ is considered constructively acceptable, but it is impredicative according to the proof-theoretic analysis of predicativity. To conclude this section, I sketch some options that a constructivist could consider to support the view that inductive definitions are after all not only constructive but also predicative.

11.4.3.1 Invariance

A first option is to link the contemporary discussion on inductive definitions directly to the original debate on predicativity. The idea is to focus on strong affinities that exist between the motivation offered today for the predicativity of inductive definitions and themes that pervade the original debate on predicativity at the beginning of the twentieth century. I only consider two points, the role of infinity in this debate and the concept of set, even if an analysis of the relevant literature suggests further significant similarities.

In Sect. 11.3, I have presented a standard characterisation of predicativity in terms of lack of vicious circularity. Poincaré also offered another characterisation of predicativity in terms of a form of invariance, which seems more suitable to capture the phenomenon of inductive definitions. According to this new characterisation of predicativity, a predicatively defined set cannot be modified or disordered by an extension of the class of sets under consideration.[31] This characterisation of predicativity relates to the one in terms of vicious circularity as follows: if we consider an impredicative definition (in the sense of circular), it would seem to have the effect of extending or enlarging a set under consideration. Let us see this with an example. Suppose we are given an impredicative definition of as set X which refers to (e.g. universally quantifies over) a set G to which X belongs. For this definition to be meaningful, it would seem that we need first to fix the extent of the set G.

[31] See Poincaré (1909, 1912). See also Kreisel (1960, p. 378). Note that I am here interested in the main ideas underlying this notion, rather than in an exegesis of Poincaré's thought.

But then X would be a "new" element of G which therefore extends or disorders G. Poincaré's requirement of invariance of mathematical definitions aims at avoiding definitions of this kind: a predicative (in the sense of invariant) definition does not disorder a set when new elements are introduced.

Weyl (1918) proposed a detailed predicative foundation of analysis that bears important analogies with Poincaré's new characterisation of predicativity. Weyl's discussion, like Poincaré's, is characteristically bound up with a stark rejection of actual infinity in mathematics, advocating instead a potentialist view of infinity. In Weyl's case, this is further directly connected with his explicit rejection of arbitrary sets. The main idea, which is reminiscent of Poincaré, is that for a correct treatment of infinite sets, one needs predicative definitions. For Weyl this means that one defines an infinite set as the extension of some property or *relation*, which may be seen as describing a step-by-step process of formation of the set.[32] In the case of analysis, which is Weyl's main focus in "Das Kontiuum", sets of natural numbers are extensions of properties built up step-by-step from the natural numbers by *repeated application of the logical operations and a principle of iteration*, with the crucial restriction of quantification to the domain of the natural numbers. Weyl calls this process of set-formation "the mathematical process" and contrasts his predicative concept of set with the dominant concept of set. He writes (Weyl, 1918, p. 20):

> Finite sets can be described in two ways: either in *individual* terms, by exhibiting each of their elements, or in *general* terms, on the basis of a rule, i.e., by indicating properties which apply to the elements of the set and to no other objects. In the case of infinite sets, the first way is impossible (and this is the very essence of the infinite).

This brings Weyl to reject the meaningfulness of the powerset of an infinite set, including the set of all subsets of the natural numbers, as it is not amenable to a general description in terms of exhaustive rules. The open-endedness of infinite sets means that the powerset of the natural numbers should be defined by rules which specify all and only its elements. (Weyl, 1918, p. 23) writes:

> The representation of an infinite set as a "gathering" brought together by infinitely many individual arbitrary acts of selection, assembled, and then surveyed as a whole by consciousness, is nonsensical; "inexhaustibility" is essential to the infinite. [...] Therefore I contrast the concept of set and function formulated here in an exact way with the completely vague concept of function which has become canonical in analysis since Dirichlet and, together with it, the prevailing concept of set.

A remarkable aspect of Weyl's concept of set is the inductive generation of the properties of the natural numbers through iterated application of the logical operations (with restricted quantifiers). Weyl lucidly highlights the crucial role of this iteration for the mathematical process.

I have emphasised two significant aspects in Weyl's "Das Kontinuum": the objection to the powerset of an infinite set and the role of a potentialist view of infinity. It is interesting to compare these with more recent discussions on predicativity. The impredicativity and the arbitrariness of the powerset of an infinite

[32] See also Cantini (2022) for discussion.

set is also exposed in a fundamental article by Myhill (1975), in which the author sets out the details of a constructive set theory that, notwithstanding its use of intuitionistic logic, bears strong formal affinities with ZF set theory. Myhill replaces the powerset axiom of ZF with a constructively weaker axiom of exponentiation, as the first is seen as lacking constructive justification.[33] Myhill's criticism of the powerset axiom of ZF is particularly clear, and deserves quoting:

> Power set seems especially nonconstructive and impredicative compared with the other axioms [of set theory]: it does not involve, as the others do, putting together or taking apart sets that one has already constructed but rather selecting out of the totality of all sets, all those that stand in the relation of inclusion with a given set (Myhill, 1975, p. 351).

We have here the opposition between, on the one side, the arbitrariness of the powerset of an infinite set, whose justification seems to require the prior availability or even surveyability of an infinite mathematical domain, with, on the other side, the rule-like construction of a set. This strongly resonates with the typical constructive appeal of inductive definitions, which has been repeatedly stressed above: the rule-like build up of a set from some initial unproblematic elements. We also saw that inductive definitions are often introduced today to eliminate problematic uses of the powerset of infinite sets, almost as if they were computationally approximating from below as much as possible of the powerset of an infinite set. Furthermore, the monontonicity of inductive definitions would seem to ensure that at no time the generation of new elements "disrupts" or modifies earlier fragments of the set – at least in the sense that what has entered the set at a certain stage cannot leave it at subsequent stages.

As to the role of potential infinity in Weyl's analysis, this also has a counterpart in more recent discussions. In the fundamental (Lorenzen and Myhill, 1959), the authors introduce (generalised) inductive definitions and argue for their constructivity. In their conclusion they write that the method of inductive definitions

> exhausts those means of definition at present known which are acceptable from a standpoint which rejects the actual infinite.[34]

In view of the rule-like character and the monotonicity of inductive definitions, as well as these remarkable similarities with recent discussions on predicativity, it seems at least possible to give a predicative justification of these constructions along the lines of Poincarè and Weyl's considerations. The challenge here is to sharpen the notion of invariance in a way that more directly applies to the case of inductive definitions.[35]

[33] The axiom of exponentiation allows us to collect in a set all the functions from a set A to a set B. This is constructively weaker than the full powerset (Aczel, 1978; Myhill, 1975).

[34] I would like to thank a referee for drawing my attention to this passage and to Lorenzen (1958).

[35] A thorough discussion of this point would require careful consideration of Lorenzen's work. See e.g. Lorenzen (1958). Note that one could argue that the term "predicativity" is now been used to refer to a different phenomenon altogether compared with that giving rise to the Γ_0 limit. This seems to be Feferman's point of view in Feferman (1964, p. 4–5), when discussing especially

11.4.3.2 Logic

The constructivist could explore another, not necessarily disjoint, strategy, by pursuing the question whether the underlying logic has a role to play in an assessment of the predicativity of inductive definitions. More specifically, one could argue that a shift to intuitionistic logic makes a defense of the predicativity of inductive definitions more plausible.

The idea is to extend to the case of generalised inductive definitions considerations that arise for the inductive definition of the natural numbers. Elsewhere, I analyse the presuppositions of a predicativist argument for *intuitionistic logic* inspired by Dummett's argument for indefinite extensibility.[36] This argument makes essential use of a claim that mathematical induction involves a form of circularity.[37] In Sect. 11.4.1, we saw the inductive definition of the natural numbers as the least set containing 0 and closed with respect to the successor operation. The closure condition for the natural numbers is expressed by the principle of mathematical induction. Mathematical induction is a fundamental principle in arithmetic, which enables us to prove universal statements as follows: it suffices to show that a property, say F, holds of the first natural number, 0, and that it progresses from a number to the next one, i.e. that if F holds of n, it also holds of $suc(n)$. Then we can conclude that F holds of *every* natural number.

Though the natural numbers are considered unproblematic according to both forms of predicativity under examination here, one may claim that a thorough predicativist perspective ought to recognize the impredicativity of the principle of mathematical induction.[38] The worry regarding induction is that this minimality condition involves a circularity. One way of expressing this concern is by observing that the principle of mathematical induction is stated for *arbitrary* properties. Therefore, it also applies to those properties, F, that refer to the whole set of natural numbers. In other terms, the formula which describes the property F in the principle of mathematical induction may contain unrestricted number quantifiers, like, for example, a universal quantifier ranging over *all* the natural numbers. The natural numbers would then be defined in terms of the whole collection of natural numbers, against the VCP.[39]

The thought scrutinised in Crosilla (2020) is that while standard interpretations of classical quantification require the availability of each element of the domain prior to quantification over it, giving rise to the difficulties above, an intuitionistic

Lorenzen and Wang's work on predicativity. I am persuaded this is a complex issue that would require careful consideration.

[36] See Crosilla (2020).

[37] See Dummett (1963), Nelson (1986) and Parsons (1992).

[38] See e.g. Dummett (1963), Nelson (1986) and Parsons (1992).

[39] See Nelson (1986) and Parsons (1992). See also Crosilla (2016, 2020) for a detailed analysis of the natural number case.

universal quantifier, at least in some cases, may by given a generic interpretation.[40] More specifically, in an intuitionistic context the "problematic" use of universal quantification over the natural numbers that we find in the principle of mathematical induction may be given a generic interpretation. This would seem to suffice to eliminate the perceived difficulty involved with the circularity of mathematical induction. The constructivist could hope that considerations of this kind could be extended to the case of generalised inductive definitions, so to ease the difficulty with the apparent circularity of the closure condition which was discussed in Sect. 11.4.2.

11.4.3.3 Trees

In Sect. 11.4.2, we saw that a generalised inductive definition can be presented in stages, indexed by classical (impredicative) ordinals. The constructivist could employ well-founded trees within an intuistionistic context to play a role analogous to that of the classical ordinals in the classical context and argue that, constructively, well-founded trees are directly predicatively justified. This seems to be a view often put forth by constructive type theorists. For example, Palmgren (1998) compares predicativity given the natural numbers with "the *constructivist notion of predicativity* which recognises a construction as predicative if it has a clear inductive structure, e.g. W-sets and superuniverses."[41] A constructivist who wished to proceed along this route, would need to explain what grants, in an intuitionistic context, a direct justification of the induction principles that express the closure of (at least some) generalised inductive definition. Perhaps, one could proceed by analogy to the case of predicativity given the natural numbers. That form of predicativity takes the natural numbers as given, and in so doing accepts as unproblematic the principle of mathematical induction. One could perhaps make a similar move in the case of inductive definitions, claiming that the relevant form of transfinite induction is predicatively justified at least on the basis of intuitionistic logic. An argument along these lines could also employ some of the considerations from the previous subsection, as one may insist that the intuitionistic focus on proofs rather than objects could make such an assumption more acceptable. Once more, Poincaré and Weyl's philosophies of mathematics could also be source of inspiration. Both mathematicians insisted on the impossibility of giving a reduction of the principle of mathematical induction. Poincaré appealed to a form of intuition to justify it. Weyl made very clear the crucial role of the principle of iteration within his predicative foundation of analysis and in mathematics more generally. One could then explore the plausibility of an approach to predicativity which takes the inductive definition of the natural numbers as paradigmatic example of more general forms of inductive definitions that are taken as given and no further reducible.

[40] See also Linnebo (2018).

[41] See also Dybjer (2012) and Dybjer and Setzer (2003).

11.5 Conclusion

In this note, I have offered motivation, stemming from the current mathematical practice, for an investigation into the notion of predicativity and especially constructive predicativity. I have highlighted the role of predicativity in current debates, and its key role in concrete practical applications, where it acts as a criterion for the correctness of computation and for consistency. Inductive definitions represent powerful expressive means of definition, which are increasingly employed in the constructive practice. In that context, they are usually considered justified not only from a constructive but also from a predicative point of view. The predicative justification of inductive definitions, though, requires further thought. One of the points of concern is the fact that these definitions are impredicative according to the proof-theoretic analysis of predicativity put forth by Kreisel, Feferman and Schütte. More specifically, the proof-theoretic strength of theories of inductive definitions exceeds by far the proof-theoretic strength of theories which are recognised as predicative according to that analysis. This leaves open the question of what could be taken to offer predicative justification to inductive definitions from a constructive perspective and, therefore, what characterises constructive predicativity. I have offered three (non–exclusive) suggestions. One would be to explore the original debates on predicativity, especially Poincaré and Weyl's contributions, as they present us with ideas which have significant affinities with those emerging in more recent debates. Another option is to focus on the role of different understanding of quantification, and explore whether a shift to intuitionistic rather than classical logic could eliminate or alleviate the perceived difficulties with inductive definitions. Finally, the third option is to explore the role of the paradigmatic example of the natural numbers, with its principle of induction, for a new constructive route to a justification of the stratification in stages of an inductively defined set. Here the principal question is what could grant the constructivist's belief that the relevant well-founded trees are constructively and predicatively acceptable.

Acknowledgments I would like to thank the anonymous referees for helpful comments and the editors of this volume, Stefano Boscolo, Gianluigi Olivieri and Claudio Ternullo for their determination in bringing the project of this volume to completion. I am grateful to Andrea Cantini and Øystein Linnebo for comments on an earlier version of this article. The research leading to this article has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 838445.

References

Aczel, P. 1977. An introduction to inductive definitions. In *Handbook of mathematical logic*, volume 90 of *Studies in logic and the foundations of mathematics*, ed. J. Barwise, 739–782. Elsevier.

Aczel, P. 1978. The type theoretic interpretation of constructive set theory. In *Logic colloquium '77*, ed. A. MacIntyre, L. Pacholski, and J. Paris, 55–66. New York: Amsterdam.

Aczel, P. 1986. The type theoretic interpretation of constructive set theory: Inductive definitions. In *Logic, methodology, and philosophy of science VII*, ed. R.B. Marcus, G.J. Dorn, and G.J.W. Dorn, 17–49. New York: Amsterdam.

AGDA. 2020. Agda wiki. Available at http://wiki.portal.chalmers.se/agda/pmwiki.php.

Barbanera, F., and S. Berardi. 1996. Proof-irrelevance out of excluded-middle and choice in the calculus of constructions. *Journal of Functional Programming* 6(3): 519–525.

Barwise, J. 1975. *Admissible sets and structures. An approach to definability theory*. Berlin: Springer.

Beeson, M. 1985. *Foundations of constructive mathematics*. Berlin: Springer.

Benacerraf, P., and H. Putnam. 1983. *Philosophy of mathematics: Selected readings*. Cambridge University Press.

Bishop, E. 1967. *Foundations of constructive analysis*. New York: McGraw-Hill.

Bridges, D.S., and F. Richman. 1987. *Varieties of constructive mathematics*. Cambridge University Press.

Buchholz, W., S. Feferman, W. Pohlers, and W. Sieg. 1981. *Iterated inductive definitions and subsystems of analysis*. Berlin: Springer.

Cantini, A. 2022. Truth and the philosophy of mathematics. This volume.

Constable, R.L., and et al. 1986. *Implementing mathematics with the nuprl proof development system*. Englewood Cliffs: Prentice–Hall.

Coquand, T. 1989. Metamathematical investigations of a calculus of constructions. Technical report, INRIA.

Coquand, T., and G. Huet. 1986. The calculus of constructions. Technical Report RR-0530, INRIA.

Coquand, T., G. Sambin, J. Smith, and S. Valentini. 2003. Inductively generated formal topologies. *Annals of Pure and Applied Logic* 124(1): 71–106.

Crosilla, L. 2016. *Constructivity and Predicativity: Philosophical foundations*. Ph. D. thesis, School of Philosophy, Religion and the History of Science, University of Leeds.

Crosilla, L. 2017. Predicativity and Feferman. In *Feferman on foundations: Logic, mathematics, philosophy*, Outstanding contributions to logic, ed. G. Jäger and W. Sieg. Springer. Forthcoming.

Crosilla, L. 2019. The entanglement of logic and set theory, constructively. *Inquiry* 0(0), 1–22.

Crosilla, L. 2020. From predicativity to intuitionistic mathematics, via Dummett. Unpublished Manuscript.

Dean, W., and S. Walsh 2016. The prehistory of the subsystems of second-order arithmetic. *Review of Symbolic Logic* 10: 357–396.

Dummett, M. 1963. The Philosophical Significance of Gödel's Theorem. *Ratio* 5: 140–155.

Dybjer, P. 2000. A general formulation of simultaneous inductive-recursive definitions in type theory. *The Journal of Symbolic Logic* 65(2): 525–549.

Dybjer, P. 2012. Program testing and the meaning explanations of Martin-Löf type theory. In *Epistemology versus Ontology, Essays on the Philosophy and Foundations of Mathematics in Honour of Per Martin-Löf*, ed. P. Dybjer, S. Lindström, E. Palmgren, and B. Sundholm.

Dybjer, P., and A. Setzer 2003. Induction–recursion and initial algebras. *Annals of Pure and Applied Logic* 124(1): 1–47.

Feferman, S. 1964. Systems of predicative analysis. *Journal of Symbolic Logic* 29: 1–30.

Feferman, S. 1975. A language and axioms for explicit mathematics. In *Algebra and logic*, volume 450 of *Lecture notes in mathematics*, J. Crossley, 87–139. Berlin: Springer.

Feferman, S. 1988. Weyl vindicated: Das Kontinuum seventy years later. In *Temi e prospettive della logica e della scienza contemporanee*, ed. C. Cellucci and G. Sambin, 59–93.

Feferman, S. 2004. Comments on 'Predicativity as a philosophical position' by G. Hellman. *Review Internationale de Philosophie* 229(3).

Feferman, S. 2005. Predicativity. In *Handbook of the philosophy of mathematics and logic*, ed. S. Shapiro. Oxford: Oxford University Press.

Feferman, S. 2013a. The proof theory of classical and constructive inductive definitions. A forty year saga, 1968–2008. In *Ways of proof theory*, ed. R. Schindler, 7–30. De Gruyter.

Feferman, S. 2013b. Why a little bit goes a long way: Predicative foundations of analysis. Unpublished notes dating from 1977–1981, with a new introduction. Retrieved from the address: https://math.stanford.edu/~feferman/papers.html.
Friedman, H. 1973. The consistency of classical set theory relative to a set theory with intuitionistic logic. *Journal of Symbolic Logic* 38: 315–319.
Girard, J. 1972. *Interprétation fonctionnelle et élimination des coupures de l'arithmétique d'ordre supérieur*. Ph. D. thesis, These d'Etat, Paris VII.
Gödel, K. 1944. Russell's mathematical logic. In *The philosophy of Bertrand Russell*, ed. P.A. Schlipp, 123–153. Northwestern University, Evanston and Chicago. Reprinted in Benacerraf and Putnam (1983). (Page references are to the reprinting).
Gonthier, G. 2008. Formal proof–the four-color theorem. *Notices of the American Mathematical Society* 11(55): 1382–1393.
Kreisel, G. 1958. Ordinal logics and the characterization of informal concepts of proof. In *Proceedings of the International Congress of Mathematicians (August 1958)*, 289–299. Paris: Gauthier–Villars.
Kreisel, G. 1960. La prédicativité. *Bulletin de la Societé Mathématique de France* 88: 371–391.
Linnebo, O. 2018. Generality explained. Unpublished manuscript.
Lorenzen, P. 1958. Logical reflection and formalism. *The Journal of Symbolic Logic* 23(3): 241–249.
Lorenzen, P., and J. Myhill. 1959. Constructive definition of certain analytic sets of numbers. *Journal of Symbolic Logic* 24: 37–49.
Martin-Löf, P. 1975. An intuitionistic theory of types: Predicative part. In *Logic Colloquium 1973*, ed. H.E. Rose and J.C. Shepherdson. Amsterdam: North–Holland.
Martin-Löf, P. 1982. Constructive mathematics and computer programming. In *Logic, methodology, and philosophy of science VI*, ed. L.J. Choen. Amsterdam: North–Holland.
Martin-Löf, P. 1984. *Intuitionistic type theory*. Naples: Bibliopolis.
Martin-Löf, P. 2008. The Hilbert–Brouwer controversy resolved? In *One hundred years of intuitionism (1907 – 2007)*, ed. E.A. van Atten, 243–256. Publications des Archives Henri Poincaré .
Moschovakis, Y. 1974. *Elementary induction on abstract structures (Studies in logic and the foundations of mathematics)*. American Elsevier Pub. Co.
Myhill, J. 1975. Constructive set theory. *Journal of Symbolic Logic* 40: 347–382.
Nelson, E. 1986. *Predicative arithmetic*. Princeton: Princeton University Press.
Nordström, B., K. Petersson, and J.M. Smith. 1990. *Programming in Martin-Löf's type theory: An introduction*. Clarendon Press.
Palmgren, E. 1992. Type-theoretic interpretation of iterated, strictly positive inductive definitions. *Arch Math Logic* 32: 75–99.
Palmgren, E. 1998. On universes in type theory. In *Twenty–five years of type theory*, ed. G. Sambin and J. Smith. Oxford: Oxford University Press.
Parsons, C. 1992. The impredicativity of induction. In *Proof, logic, and formalization*, ed. M. Detlefsen, 139–161. London: Routledge.
Poincaré, H. 1905. Les mathématiques et la logique. *Revue de Métaphysique et Morale* 1: 815–835.
Poincaré, H. 1906a. Les mathématiques et la logique. *Revue de Métaphysique et de Morale* 2: 17–34.
Poincaré, H. 1906b. Les mathématiques et la logique. *Revue de Métaphysique et de Morale* 14: 294–317.
Poincaré, H. 1909. La logique de l'infini. *Revue de Métaphysique et Morale* 17: 461–482.
Poincaré, H. 1912. La logique de l'infini. *Scientia* 12: 1–11.
Rathjen, M. 2005. The constructive Hilbert program and the limits of Martin–Löf type theory. *Synthese* 147: 81–120.
Rathjen, M., E. Griffor, and E. Palmgren. 1998. Inaccessibility in constructive set theory and type theory. *Annals of Pure and Applied Logic* 94: 181–200.
Russell, B. 1906a. Les paradoxes de la logique. *Revue de métaphysique et de morale* 14: 627–650.

Russell, B. 1906b. On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society* 4: 29–53.
Russell, B. 1908. Mathematical logic as based on the theory of types. *American Journal of Mathematics* 30: 222–262.
Russell, B. 1973. *Essays in analysis*, ed. D. Lackey. New York: George Braziller.
Sambin, G. 1987. Intuitionistic formal spaces – a first communication. In *Mathematical logic and its applications*, ed. D. Skordev, 187–204. Plenum.
Schütte, K. 1965a. Eine Grenze für die Beweisbarkeit der Transfiniten Induktion in der verzweigten Typenlogik. *Archiv für mathematische Logik und Grundlagenforschung* 7: 45–60.
Schütte, K. 1965b. Predicative well–orderings. In *Formal systems and recursive functions*, ed. J. Crossley and M. Dummett. North–Holland, Amsterdam.
Simpson, S.G. 1988. Partial realizations of Hilbert's program. *Journal of Symbolic Logic* 53(2): 349–363.
Simpson, S.G. 1999. *Subsystems of second order arithmetic*. Perspectives in Mathematical Logic. Springer.
The Coq Development Team. 2020. Coq. https://coq.inria.fr.
Troelstra, A.S. 1999. From constructivism to computer science. *Theoretical Computer Science* 211: 233–252.
Univalent Foundations Program, T. 2013. *Homotopy type theory: Univalent foundations of mathematics*. Institute of Advanced Studies.
Wang, H. 1959. Ordinal numbers and predicative set theory. *Zeitschr. f. math.Logik und Grundlagen d. Math.* 5: 216–239.
Weyl, H. 1918. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*. Veit, Leipzig.
Whitehead, A.N., and B. Russell. (1910, 1912, 1913). *Principia mathematica, 3 Vols.*, Vol. 1. Cambridge: Cambridge University Press. Second edition, 1925 (Vol 1), 1927 (Vols 2, 3); abridged as Principia Mathematica to *56, Cambridge: Cambridge University Press, 1962.

Chapter 12
Truth and the Philosophy of Mathematics

Andrea Cantini

Abstract Is truth – *qua* a primitive notion – fit to play an independent role in the philosophy of mathematics and in the foundational investigations? The problem is handled by surveying axiomatic theories of truth and their implications, with a main concern for ontological and epistemological issues.

Keywords Axiomatic theories of truth · Typed theory · Kripke-Feferman theory

2010 Mathematics Subject Classification: 03F03, 03F25, 03F35, 3F40, 03A05.

12.1 Introduction

The paper intends to be a complement to Cantini (2017) – only from the author's present perspective, and with a different focus and a wider scope.[1]

On one side, truth is regarded as a tool for *ontological reduction*. On the other side, truth is naturally enjoined when handling the classical epistemological problem of *implicit commitment:* what ought we to accept once we have made a commitment to a mathematical system S?

A warning: we stick to a view of the philosophy of mathematics,[2] which follows the pattern of logic, so most successful with *analysing the language of mathematics and its verificational methods*, the logical analysis of the structure of mathematics and the corresponding focus on formal systems as objects of study.

Research partially supported by PRIN 2017.

[1] Cantini (2017) being dedicated mainly to Feferman's work.

[2] See Feferman (1998a) pp.123–24.

A. Cantini (✉)
DILEF, Università di Firenze, Firenze, Italy
e-mail: andrea.cantini@unifi.it

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_12

We do not pretend to offer a wider look of mathematical knowledge as articulated e.g. in the heuristic view of Cellucci (2017) (recall the opposition *logic of mathematical discovery* vs. *logical structure of mathematics*). In this respect we send the reader to the paper *Working foundations–'91* in Feferman (1998a) and to p.92. Accordingly, as to the contents, in Sect. 12.2 we briefly introduce the issue of truth in the context of mathematical philosophy. Section 12.3 considers philosophical uses of truth: in particular truth as a means for the foundation of classes and sets. This step opens up the scenario of a conceptualistic approach, where the distinction between predicative and impredicative definitions plays a role.

The conceptualistic approach leads to the technical consideration of theories of the truth predicate. The minimal one is the theory of compositional truth CT. In Sect. 12.4 we describe CT and state some fundamental results: conservation and speed-up. There is an attempt to assess the philosophical meaning of CT with respect to *ontological reduction*. As to this issue, we consider an essentially negative result concerning the attempt of outlining a definitionist foundation of mathematics, with reference to the issue of simulating power set in intensional context.

We discuss the issue of reflective expansion of concepts and principles (see Feferman (1998a), p.120). This route starts with iterating typed truth and hence expanding CT to hierarchical theories of truth. We then proceed in Sect. 12.5 to reflective closure; this justifies introducing KF and eventually standard predicativity. It also opens up the route towards metapredicativity in the sense of the Bern school (e.g. see Jäger et al. 1999).

12.2 Truth in the Philosophy of Mathematics

Since Plato – and in general from a naive realistic naive point of view – truth in mathematics is understood as pointing to an abstract ideal world, and hence the main problem – metaphysical as well epistemological – is how to make sense of, and how to gain access to the abstract universe of mathematical objects. Just to refresh the context in terms that may be familiar from classical references (see Benaceraff's arguments), if our best theory of knowledge is based on causal reliable interaction between mathematical objects and the mathematicians as knowing subjects, how is this possible? One way out is to follow a standard route, according to which mathematical objects live or come into existence within suitable *structures*, which are assumed to exist in connection with a certain body of (consistent) knowledge – typically embodied in a mathematical theory; in turn, these structures are intended to verify the statements of the given theory and its logical consequences.

The object theory under analysis may not be *complete*, our knowledge can be *partial* with respect to certain statements, that are left undetermined as to their truth value. But accessibility is granted by the fact that *we are able to display our*

notions in a systematic way in a web of notions, by building concepts and proof constructions which are sound with respect to the given principles.[3]

Of course, under a close scrutiny, we may like to stress the epistemological aspects, i.e. proofs and definitions: but this attitude would apparently drive us to a shift from *the truth of a statement* to the *the grounds of an assertion*, and hence to emphasize the relationship between truth and its *justification*, e.g. provability, which is a reason for a substantial recasting of our understanding of logic and mathematics. Just to simplify our life, we here assume a sort of *epochè* concerning all the epistemological aspects and we simply do not discuss truth in the context of, say, a theory of types in the sense of Martin-Löf, or even in a classical sense, as in Tait (1983, 1986).

We rather conform to the neutral idea that the concept of truth is central and indispensable: in order to understand the meaning of a statement, you must grasp the concept of truth for the given language as primitive.[4] The concept of truth and the concept of meaning are inextricably linked (see Dummett 2004): typically, if we grasp the meanings of two statements and one can be accepted without the other being true, they must have different meaning. Meaning and truth must be simultaneously explained. Technically, this idea is susceptible of a precise formalization, in such a way that the notion of truth and proposition – as content of a statement – are simultaneously given e.g as in Aczel (1980) or Feferman (2008).

On the other hand, we shall stick to the naive intuition of *semantic ascent* and *semantic descent*: the sentences A and "A is true"– in symbols henceforth $T(\ulcorner A \urcorner)$ – have the same content and hence are intersubstitutable. Axiomatically, this intuition is embodied by the well-known disquotational conception of truth, embodied by the T-schema: for arbitrary sentence A

$$T(\ulcorner A \urcorner) \leftrightarrow A \tag{12.2.1}$$

In a certain sense the disquotational conception of truth seems simpler and even the more basic (see Horsten and Leigh 2017). But (12.2.1) is a well-known source

[3] The model we have in mind derives from practice, it is simply either the set of informal elucidations that precede the informal presentation or development of a theory, axiomatic or not.

[4] Let me cite Tarski himself, Tarski (1944), p.352: When a language is unable to define truth, we then have to include the term "true" or some other semantic term, in the list of undefined terms of the meta-language, and to express fundamental properties of the notion of truth in a series of axioms. There is nothing essentially wrong in such an axiomatic procedure, and it may prove useful for various purposes.

In Tarski (1956), p.266, Tarski similarly writes that for some of the languages for which truth cannot be defined, we can nevertheless make "consistent and correct use" of the concept of truth by way of taking truth as a primitive notion, and giving it content by introducing the relevant sorts of axioms.

of difficulties and this forces us to look at the present debate on the semantical foundations from a different point of view, that of the philosophy of mathematics.[5]

However, this is not our main concern. Instead we like to explore the role of truth predicate in itself, as a primitive notion, and we rephrase the initial problem as:

(*) can the *truth predicate* play an independent role in the philosophy of mathematics and in the foundational investigations?

All these themes will be dealt with in the light of recent technical developments and the current attempts to recast in a new light forms of predicativism and conceptualism.[6]

12.3 Foundational Uses of Truth Predicates

A straightforward positive answer to (*), which naturally comes to the mind, is simply that theories of comprehension and satisfaction are closely related: a is an element of the class $\{x : \varphi(x)\}$ for a formula $\varphi(x)$ roughly amounts to stating that the formula $\varphi(x)$ is satisfied by a or that the formula $\varphi(x)$ is true of a or, finally, that the predicate defined by P *truly applies* to a.

Using this observation, one can reduce class theories to theories of satisfaction or truth; furthermore it seems that we have an *ontological* gain: *one can replace quantification over classes by quantification over formulas* or propositional functions (whatever they are). The basic ideas are to be credited at least to Russell's no-classes theory; the source is Russell (1906), pp.45–47, but see (*Principia Mathematica*, section 20*), and also *Introduction to Mathematical Philosophy*, chap. 13 (Russell 1919).[7]

But *the solution is problematic*. First of all, due to the antinomies, the concept of truth ought to be considered with suspicion as well, and the reduction of mathematical theories to theories of satisfaction or truth does not look too attractive. Nevertheless – thinking of Russell's emphasis on the method of logical construction (*the honest toil* mentioned in Russell (1919), page 71) – it might be more attractive a choice. But is this a sufficient reason? As stated elsewhere (Cantini, 2017), 288, we apparently "want to replace a respectable mathematical theory with a 'philosophical' theory of truth or satisfaction, because of a possible reduction of the ontological commitment to sets to a lighter(?) ideological commitment to notions such as truth." Instead of relying on Russellian philosophy, let us briefly explore an

[5] The link between semantical investigations and foundations of mathematics is certainly not new, see the work following Kripke (1975) by Feferman and others (as documented in the references of Feferman (2008)) and all the recent investigations of axiomatic theories of truth (Cantini, 2017).

[6] We have in mind constructive set theory as developed by Aczel and Rathjen (2001), explicit mathematics à la (Feferman, 1998c), metapredicativity in the sense of the Bern school (see Jäger 2005).

[7] But consider the approach in Parsons (1971, 1974), Cantini (1996), and recently Schindler (2018).

alternative answer to (*) and a different reason for choosing truth. This is simply that *we like to stand by Weyl*: instead of making sense of sets as a domain of entities closed under certain operations and postulates, we might be willing to make sense of sets by means of definition as specifications[8] and to stick to a view inspired by a *conceptualistic approach*.

Let me for a moment dwell on it. The fact is that this is problematic, too. As stated by Parsons (2002), 378,

> Hilbert views both Russell and Weyl as seeking to reduce the concept of set to predication.

But then Hilbert's criticism is that

> ... we must ask what *there is a predicate P* should mean. In axiomatic set theory the quantifier *there is* always relates to the underlying domain B. In logic we can, to be sure, think of predicates as collected to a domain, but this domain of predicates *cannot in this case be considered as something given at the outset*, the predicates must be constructed by logical operations and the domain of predicates is determind *only afterward* by the rules of logical construction. From this it is clear that in the rules of logical construction of predicates references to the domain of predicates cannot be allowed. For otherwise a *circulus vitiosus* would arise.[9]

While sets are assumed to form a domain of *completed totalities*, underlying basic closure conditions and operations, and hence no vicious circularity should be involved in quantifying over sets, on the contrary predicates come into being via logical operations and form an *open totality*. There might be room for impredicative definitions at the level of the given *Operationsbereich* (individuals and operations there upon are given).

Predicates are generated via logical operations and hence, according to Hilbert, the domain of predicates depends upon the rules of logical construction: by means of these rules, we can refer to predicates only at the price of a vicious circle (see Hilbert, p.31, *Probleme der mathematischen Logik*, SS 1920), or unless we have established some sort of logical *order* or *hierarchy*.

After all, how seriously can we choose truth as a notion for grounding sets? A natural (minimal) answer – in the spirit of Weyl (1910) – is that there are finitely many construction principles for sets corresponding to the *elementary* logical operations – Boolean operations, projections and combinatorial operations[10]– and handled by the very truth predicate. In an extensional vein, there is a correspondence between natural predicative non-vicious set existence principles and the logical conditions grounding the biconditionals of the form (12.2.1) for elementary formulas. Do the logical principles deserve philosophical priority over set principles? This is only a start, but it opens up a direct confrontation with most recent research.

The systems we are going to briefly touch upon below are listed in increasing proof-theoretic strength: they range from systems of arithmetical strength to systems

[8] See Weyl (1910).

[9] We quote Hilbert's text as translated in Parsons (2002), p.378; see also p.74, Mancosu (1998), and see Hilbert (2013).

[10] E.g. permutations, duplications, identification, expansion of variables.

higher up. There are not only rich models for non-trivial notions of truth, but also well-understood systems – classical as well as constructive – that can be fruitfully compared as to their significance in the foundational investigations. For the sake of space, we only explicitly consider: the compositional typed theory of truth CT; the type-free theories KF and VF.[11] Incidentally the classification results show that it is not true that the theories of truth lack 'deductive power', and that theories of truth – be they axiomatized by typed or untyped T-sentences – compare well with mathematical theories with comprehension axioms. The real issue is on the conceptual priority and this will be left open to discussion in the present paper.

A philosophical general point concerns *ontological reduction*: as written in (Halbach 2011, 331) *even the commitment to sets that are not predicatively definable can be replaced with strong semantic commitments*. But it is to be seen whether strong ontological assumptions can be reduced to semantic assumptions. In significant cases, the ontology of truth theories is typically given by a countable set of individual objects – usually taken as the natural numbers. On the other hand, proof theoretic strength is often the consequence of set existence. But it is also true that the commitment to *sets that are not predicatively definable* can be obtained via semantic principles (see VF).

Of course the value of such a reduction is debatable and lies at the core of the philosophical problem: is it preferable to rely upon the existence of an ideal structure – a universe of sets and operations – or a truth predicate?

Remark 1 More on Parsons's views (see Parsons (1971, 1974)). The issue of the relationship between predicative class theories and theories with satisfaction predicate is investigated by Parsons (1974), showing the mutual interpretability of predicative class theory and a weak theory of satisfaction (with the satisfaction predicate not allowed in the axiom schemata such as replacement and induction). Parsons argues that a satisfaction theory seems inherently predicative since it refers explicitly only to the formulae of the given language; hence he discusses the difficulties of identifying sets with extensions of predicates; he favours the view that set existence principles are not reducible to what the extension of predicates would simply suggest.

The predicativist view of classes has also been closely investigated in the recent essay: *Predicativism about Classes* by Fujimoto (2019).

12.4 Compositional Truth

First of all, we deal with the compositional theory CT. This system is deceivingly simple, but utterly non-trivial both on the semantical side as well from the point of view of proof theory. We follow essentially the standard notations and the exposition

[11] For all these theories, we refer to the monograph (Halbach, 2011).

in Halbach (2011). We start with the language of Peano Arithmetic PA, $\mathcal{L}_{\mathsf{PA}}$ with a unary predicate T for truth T.

(i). Quantifiers $\forall s$ and $\forall t$ range over the codes of $\mathcal{L}_{\mathsf{PA}}$-terms; $\forall t$ is short for $\forall x\,(\mathrm{ClTerm}(x) \rightarrow \ldots)$, where $\mathrm{ClTerm}(x)$ means that x is a closed $\mathcal{L}_{\mathsf{PA}}$-term; $^\circ$ represents in PA a recursive function taking a code of a closed $\mathcal{L}_{\mathsf{PA}}$-term and returning its numerical value;
(ii). $\mathrm{Sent}_{\mathsf{PA}}(x) := x$ is a code of a sentence of the language of PA. $\dot{\neg}$ represents in PA negation; e.g., $\dot{\neg}\ulcorner 0 = 0\urcorner = \ulcorner 0 \neq 0\urcorner$.
(iii). $\dot{\wedge}$ represents conjunction in PA.
(iv). $\dot{\forall}$ represents a function that takes a code of a variable, a code of a formula and returns the universal quantification of the formula with respect to the variable.
(v). $x[y/z] :=$ the code of the result of substituting a term encoded by y for a variable encoded by z in a formula encoded by x.
(vi). $\mathrm{Prov}_{\mathsf{S}}(x)$ represents formal provability in a theory S; so $\mathrm{Prov}_{\mathsf{PA}}(x)$ represents formal provability in PA.

Finally, $\mathsf{GRF}_{\mathsf{S}}$, the *global reflection axiom* over the theory S, is the sentence

$$\forall x(\mathrm{Prov}_{\mathsf{S}}(x) \rightarrow T(x)); \tag{12.4.1}$$

the uniform reflection schema for S $\mathsf{URF}_{\mathsf{S}}$ has the form

$$\forall x(\mathrm{Prov}_{\mathsf{S}}(\ulcorner A(\dot{x})\urcorner) \rightarrow A(x)), \tag{12.4.2}$$

where $A(x)$ is an arbitrary formula of S.

12.4.1 CT-*Axioms*

Peano arithmetic PA is our base theory and for PA the compositional axioms can be chosen as follows:

(i) $\forall s\,\forall t\,\big(T(s\doteq t) \leftrightarrow s^\circ = t^\circ\big)$ and similarly for other predicates other than $=$, except for the special predicate T
(ii) $\forall x\,\big(\mathrm{Sent}_{\mathsf{PA}}(x) \rightarrow (T(\dot{\neg}x) \leftrightarrow \neg T(x))\big)$
(iii) $\forall x\,\forall y\,\big(\mathrm{Sent}_{\mathsf{PA}}(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow T(x) \wedge T(y))\big)$
(iv) $\forall v\,\forall x\,\big(\mathrm{Sent}_{\mathsf{PA}}(\dot{\forall} vx) \rightarrow (T(\dot{\forall} vx) \leftrightarrow \forall t\, T(x[t/v]))\big)$

12.4.2 *Variants*

(i) CT is PA with the axioms for compositional truth and induction schema for the *full* language which also include T;

(ii) $\mathrm{Ind}_{\mathsf{PA}}(a)$ is the formula expressing that a is the code of the universal closure of a PA-instance of induction; the *internal arithmetical induction* I-ind is:

$$\forall a(\mathrm{Ind}_{\mathsf{PA}}(a) \to T(a))$$

Modulo the compositional axioms, it is equivalent to:

$$\mathrm{For}^1_{\mathsf{PA}}(a) \wedge T(a(0)) \wedge \forall u(T(a(u)) \to T(a(u+1))) \to \forall x T(a(x))$$

where $\mathrm{For}^1_{\mathsf{PA}}(x)$ expresses the fact that x is the code of an $\mathcal{L}_{\mathsf{PA}}$-formula with exactly one free variable;

(iii) CT^- is CT with the schema of induction for *arithmetical formulas*;

(iv) $\mathsf{CT}\restriction$ is CT with the full induction schema replaced by I-Ind;

(v) CT_0 is the extension of CT^- with *induction for bounded* [12] *formulas, possibly with the truth predicate* (so CT_0 includes $\mathsf{CT}\restriction$);

(vi) CT_1 is the extension of CT_0 with induction for Π_1-formulas with the truth predicate.

12.4.3 Basic Results

Crucially, if *the induction schema is expanded to the new language with truth*, the resulting theory CT (for 'compositional truth') naturally proves the soundness of Peano arithmetic, that is, the *global reflection principle* for PA as *one* axiom:

$$\forall x \left(\mathrm{Sent}_{\mathsf{PA}}(x) \wedge \mathrm{Prov}_{\mathsf{PA}}(x) \to T(x) \right).$$

By contrast, if induction is restricted, conservation over PA holds:

Theorem 1

(i) $\mathsf{CT}\restriction$ *is conservative over* PA.

(ii) CT_1 *is not conservative over* PA.

The relevant non-trivial technical developments can be found in older papers by Kotlarski et al. (1981), Lachlan (1981), and in more recent work by Halbach (2011, p.104), Enayat and Visser (2015), Leigh (2015, p.862), Łełyk and Wcisło (2017b, p.460), Łełyk and Wcisło (2017a), and Cieśliński et al. (2017).

Let me add a couple of comments on the proofs of the conservation (i). In general, an arbitrary model of PA contains non-standard numbers which encode *syntactical* notions – terms, formulas, derivations – and it is not at all clear that there exists a

[12] Formulas generated from atomic formulas via Boolean operations and bounded quantifiers $\forall x < t \ldots$ and $\exists x < t \ldots$ with $<$ representing the standard ordering on natural numbers.

truth predicate or a satisfaction predicate that *is correct with respect to these possibly non-standard objects*, i.e. terms or formulas encoded by *infinite* numbers.

The obvious conservation proof via expansion and completeness *does not work*. Instead, one has to resort to non-trivial model-theoretic constructions as given by Kotlarski et al. (1981), which involve the so-called *recursively saturated* models (see Kaye 2005). Indeed, if a model of PA has a notion of truth satisfying the natural axioms for T, then the model is recursively saturated, and conversely (see Kotlarski et al. 1981, p.293). As an alternative proof, one may refer to the method of Enayat and Visser, which is based on compactness, elementary chains; and nonetheless it can be formalized in weak arithmetic.

A natural question concerns whether direct effective methods exist that allow to verify conservation. Indeed, finitary methods have been devised only recently by Leigh. Leigh's proof in Leigh (2015) is subtle and inspired by the idea of finite approximations of non-standard syntactical objects.

On the philosophical side, the proof has applications, to the effect that the adoption of a classical truth predicate may have an *epistemological* value, at least in principle, *with respect to the resources* involved in formal derivations.

As already made clear by Gödel's seminal paper (Gödel 1986), adopting abstract notions and principles thereof – like truth or set – can provide not only new theorems (e.g. on the consistency), but also sensible reduction in length for an infinite number of already available proofs. This idea can be made precise via the notion of *superexponential speed-up* (also known as non-elementary speed-up), that we here recall following Fischer (2014, definition 2.3) and Caldon and Ignjatovic (2005, pp.780–781).

Let S be a theory in the language of PA such that the set of theorems of PA is a subset of the theorems of S. Then S has *non-elementary speed-up* over PA iff there is a sequence $\{\varphi_i | i \in \omega\}$ of formulas provable in PA, such that:

(i). for no function f with Kalmar elementary growth rate, we have, for all $i \in \omega$, $\|\varphi_i\|_{\mathsf{PA}} < f(\|\varphi_i\|_{\mathsf{S}})$;

Here, given a theory T in the language of PA, $\|\varphi\|_{\mathsf{T}}$ is the minimal n, such that φ has a formal derivation d in a theory T with at most of length n; 'length' means 'symbol-length', i.e. the number counting all the symbols occurring in d (see Fischer 2014, p.321).

In the crucial case of consistency statements, one has in general that, if Cons(PA) is the standard formalization of the consistency of PA, PA $+$ Cons(PA) has super-recursive speed-up over PA by the main theorem of Ehrenfeucht and Mycielski (2020, p.367).

It is known by Fischer (2014), corollary 6.2, that CT has a non-elementary speed-up with respect to PA. On the other hand, again by Leigh (2015), corollary 6.3, the theory $\mathsf{CT}\restriction$ which includes $\forall x(\mathrm{Ind}_{\mathsf{PA}}(x) \rightarrow T(x))$ has a speed-up over PA between exponential and superexponential.

This result should be compared with the case of the corresponding disquotational theory extending PA based on the uniform T-biconditional schema UTB_0, i.e. the schema $T(\ulcorner A(\dot{x})\urcorner) \leftrightarrow A(x)$ ($A(x)$ being an arbitrary formula of $\mathcal{L}_{\mathsf{PA}}$). It turns out

that the system based on UTB_0 has *no significant speed-up* over PA[13] and hence it is much more attuned with a deflationist account of the notion of truth.

Lastly, as kindly suggested by one of the referees, the results of Fischer and Leigh have been improved in Enayat et al. (2020): CT^-, as well as KF^- (see below), have no more than polynomial speed-up over PA. The result is implied by the fact that CT^- and KF^- are *feasibly reducible* to PA, in the sense that there is a polynomial time computable function f such that, for every proof π of an arithmetical sentence A in $\mathsf{CT}^-(\mathsf{KF}^-)$, then $f(\pi)$ is a proof in PA of the same formula A.

Remark 2 The quantitative results on speed-up suggest the adoption of forms of *mathematical instrumentalism*, according to which certain mathematical notions and theories can be regarded as technical means for facilitating proofs of statements in given accepted theoretical frameworks, e.g. PA: the idea is that expansions of the ground basic system PA are, so to speak, *instruments for making proofs shorter* and hence *easier to grasp* (Caldon and Ignjatovic 2005). This raises the problem: is instrumentalism on the same par as deflationism for a semantical theory?

12.4.4 Assessing the Value of CT*: Ontological Reduction?*

Let us now go back to the initial question (*) and try it again in the special form: What is the value of CT for the philosophy of mathematics, if any? We explicitly discuss two possible directions.

Firstly, as already hinted, we can qualify the choice of CT in terms of *ontological reduction*.

In fact, it is well-known how to prove that the theory ACA of second order arithmetic as based on arithmetical comprehension and full induction on numbers is relatively interpretable into CT: the range for first-order variables being fixed, second oder variables – intended to range over subsets of natural numbers – are interpreted as ranging over monadic formulas in the pure arithmetical language, while membership $x \in X$ is translated into $T(a(x))$, where a encodes any monadic formula $A(v)$ and $a(x)$ essentially corresponds to the arithmetical expression defining the code of the formula obtained by replacing v via in A by means of the code of the xth-numeral. Instead of definable sets – indeed predicatively definable ones at the simplest level, once quantification on natural numbers is accepted – we can fully resort to accepting an elementary theory of (arithmetical) truth. This implies that, instead of an ontology with sets, one takes a position where essentially *only natural numbers* are assumed to exist. More generally, this makes sense if one

[13] See Theorem 3.2 in Fischer (2014).

assumes that every mathematical object is *represented by a definition, following a definitionist inspiration going back to Poincarè and Weyl.*[14]

Under this reading, even a *nominalistic understanding of power set* becomes conceivable, and one can explore the possibility of avoiding impredicativity under a nominalistic intensional rendering of power set using truth and following a strictly intensional route. Assume that *X is a class* means X is a one-variable formula, that membership *a in X* corresponds to '*a makes X true*, while *X is a subset of Z* means *every a making X true makes Z true. too*. Is then consistent to assume the existence of a class $Pow^+(X)$ such that it consists of exactly those objects that encode subclasses of X under the given understanding of what a class is? Formally, assume that $u = X$ is a well-formed formula saying that u is a class Y. Then the corresponding set existence takes the form

$$\forall u(u \in Pow^+(X) \leftrightarrow \exists Y(Y = u \wedge Y \subseteq X)).$$

Here extensional quantification over subsets is replaced by means of quantification over certain individuals representing subsets as defined by monadic formulas which make sense via the truth predicate. By Jäger (1997) and Cantini (1996) the answer is negative: the existence of $Pow^+(X)$ is refuted, as soon as set existence admits elementary logical operations.[15] Hence the existence of the strong power set is inconsistent with elementary comprehension and hence there is no hope to model it in $\mathsf{CT}\restriction$. Thus predication and an intensional outlook are not viable (compare with Hilbert's quotation of the Sect. 12.3).

Alternatively, one can consider the weak power set operation Pow:

$$\forall u(u \in Pow(X) \rightarrow \exists Y(Y = u \wedge Y \subseteq X)) \wedge \tag{12.4.3}$$

$$\forall Y(Y \subseteq X \rightarrow \exists y \exists Z(y \in Pow(X) \wedge y = Z \wedge Z =_e Y)), \tag{12.4.4}$$

where $Z =_e Y$ means that Z and Y are extensionally equal (have the same elements). This is consistent even with the assumption that the class of all classes exists, but it is *inconsistent* with the so called join axiom (existence of disjoint sum), by simple diagonalization argument.

Incidentally, the existence of the so-called weak power classes is very weak proof-theoretically; it is known that adding weak power set axiom to elementary comprehension is conservative over PA (Glass 1996).

As stated in Jäger (1997), neither the strong nor the weak power set axiom seem to provide a convincing approach to sets or power types in foundational frameworks that take inspiration from *definitionism* (like explicit mathematics). Therefore there remains the question whether there is an alternative, possibly intermediate form,

[14] Its most recent version being the so-called explicit mathematics which was introduced in the Mid Seventies by Feferman (1979) and is still being developed by the Bern school and others. See also Feferman (1998b) and the comments in Parsons (1971) and Cantini (2016).

[15] These are neatly analyzed probably for the first time by Weyl (1910).

which is more interesting. Hilbert's criticism of both Russell and Weyl as seeking to reduce the concept of set to predication (see Sect. 12.3) is in a sense not neutralized by passing to a strict definitionist and intensional look.

12.4.5 CT *and Beyond*

We can justify the choice of CT if we naturally reconsider a classical theme that has been forcefully argued for – most notably by Kreisel (1970) and Feferman in the Sixties and early Seventies – namely that *the acceptance of a theory* S *enjoins the implicit acceptance of the soundness of* S. This means that the acceptance of S enjoins the acceptance of (all instances of) a *reflection principle*. If we try to answer

> What is implicit in accepting a mathematical system S? And what ought we to accept once we have made a commitment to S?

a canonical reply apparently is that we must accept a system S as true, which means to justify or accept the statement of the reflection principle *all (closed) theorems of* S *are true*. But this sentence cannot be formulated in S itself by means of a single statement; for this desideratum requires a *truth definition* T for the language of S and such a truth definition doesn't exist because of Tarski's theorem on the undefinability of truth.

Now this is the right place to come up with CT; the theory, as based on a *primitive predicate for truth.* Indeed, once CT is accepted and we systematically follow this path, we remain entangled with *hierarchical theories of truth*: for then we are committed to the soundness of CT, and thus to a truth predicate T for CT-soundness. To this end, one can add a truth predicate T_1 that applies to all sentences with T. T_1 is then axiomatized in the same way as T except that T_1 is treated as a non-special predicate symbol. Moreover quantification over sentences of the arithmetical language is replaced with quantification over sentences of the expanded arithmetical language. This procedure can be iterated and an axiomatization of Tarski's hierarchy of languages is obtained.

The crucial point is that we have to cope with a *never ending implicit commitment*, because the implicit commitment in the acceptance of PA is not exhausted by CT and the reflection principles: the iteration procedure further continues, and all this can be formally made precise by means of standard technicalities of recursion theory (Halbach 2011). *Eventually, the resulting formal framework is a standard way of representing predicativity over natural numbers!* And predicativity ought to be considered as the conception that makes precise the answer to the question: what is implicit in assuming the structure $\mathbb{N}$ of natural numbers and the general principle of induction over $\mathbb{N}$.

There are – however – objections that can be raised to the whole idea.

First of all, on the conceptual level, it is to be seen whether these iterated theories of truth whose purpose is to make explicit assumptions implicit in the acceptance of PA, are *truly implicit* in the acceptance of PA any way, given the complex ordinal

structures that are necessary to make the informal ideas precise. It is hard to accept that all the heavy machinery might be naturally ascribed to the simple acceptance of PA. Since PA proves the transfinite induction schema for any initial segment of the standard wellordering of type ε_0, the truth predicates can then be iterated up to that point. The new theory with transfinitely many truth predicates, however, proves transfinite induction for longer wellorderings. Hence the truth theories are iterated even further, following a well-known bootstrapping procedure, until a point Γ_0, i.e. the so-called Feferman–Schütte ordinal, is reached. Now the iterated truth theories very much resemble *the systems of predicative analysis*, which had been studied thoroughly in the 1960s, and Γ_0 is such that the transfinite induction principle along any $\alpha < \Gamma_0$ is regarded as predicatively acceptable ordinals.

Incidentally, we recall that there is a jump of logical order in the concepts involved. Technically, the problem is that the orderings involved have to be well-founded; but this very notions leads to second order concepts and to the delicate issue: how far can the process of reflection be iterated in a way that the proviso of strict predicativity are met.

Lastly, there might be some worries about the genuine epistemological value that can be ascribed to the soundness theorem – formalized via reflection. We have chosen the intended semantics so that axioms of PA become true, as we believe in their truth. In view of incompleteness, the important fact is that soundness –trivial as it is – is unprovable as it implies consistency. And indeed Girard (1987), p. 64 observes that there is no addition to the fact that we accept principles and rules, but maybe all this has an interest because of formal uses.

To sum up, the results on iterated truth theories are formally not extremely exciting and still have some problematic aspects that have been recently considered in the philosophical literature.

Nevertheless, the whole matter can also be reconsidered from a Gödelian point of view. By incompleteness we are doomed to search for new axioms and there is the need for introducing new axioms in order to settle undecided propositions. How to proceed? A natural procedure is just by reflection and the desideratum would be *to concentrate on the consideration of axioms which are supposed to be 'exactly as evident' as those already accepted.*[16]

It remains to be seen how to implement these ideas of reflective expansion and this will be done in the next Sect. 12.5.

12.5 Untyped Truth

From a foundational point of view, *the iterated classical truth theories are significant*: a way of carrying out the programme of determining the *reflective closure* of PA, *that is, of characterizing the theory that makes explicit what is implicit*

[16] Note that this excludes the principles about very large cardinals, determinacy, etc.

in the acceptance of PA. But the formulation of the systems of iterated truth is technically awkward and highly specific to PA. Hence the question whether it is possible to characterize the reflective closure of theories in a more elegant and applicable way, and in a way that it is independent on 'natural' ordinal notation systems and arithmetization.

It is possible, but it does require a conceptual turn. Let us go back to fundamentals: assume a generalized constructive view of mathematical knowledge where truth is associated with a semantical evaluation schema, to be read as a set of possibly infinitary inferences. Then it turns out that truth must be partial, approximate, inductively generated, and it may be simply undecided under a given schema whether a given statement has a determinate truth value. Predicates are better seen as potentially given through the evaluation process and they can be open-ended, and naturally seen as *partial operations*. This in opposition to sets, whose membership is *total*. Of course, there are predicates P that are completed, i.e. there is a stage such that, given any individual element a of the ground universe, either *it is verified* that a falls under P, i.e. $P(a)$ is verified or *it is verified that it does not*, i.e. the statement $P(a)$ is *falsified*. Partiality is intrinsic to the ontological choice; for it may happen that neither $P(a)$ is verified nor $P(a)$ is falsified.

According to this view, true statements and false ones are simultaneously generated from the basic *non-semantical facts* - true and false atomic sentences. All this gives rise to a sort of indefinite monotone *investigative process* –just to use Feferman's words–, which works in stages – classically ordinals – and the ordinals emerge from the theory itself and they are not imposed on it from the outside. Furthermore, truth is self-applicable, no need of an explicit hierarchical structure.

From the point of view of foundations, the technical achievements can be ascribed to the application of generalized recursion theory, and definability theory in a general setting which fully recovers tools from that part of the set theoretic tradition, which is linked with the attempt of some sort of constructive understanding: the semintuitionistic tradition having its source in Poincaré, Baire, Borel and Lebesgue (see Cantini (1985b) and most recent work due to Rathjen (2016)).

12.5.1 Kripke–Feferman

If we try to make these ideas formally precise, we are naturally led to formalize the clauses of Kleene strong three valued semantical schema and hence to the well-known KF-theory (over PA). For the reader's sake, let us summarize a few points about KF.

The axioms of KF comprise PA, full induction schema, and

(i). $\forall s\, \forall t \left(\left(T(s \doteq t) \leftrightarrow s^\circ = t^\circ \right) \wedge \left(F(s \doteq t) \leftrightarrow s^\circ \neq t^\circ \right) \right)$, and similarly for other predicates other than $=$, except for the special predicates T and F;

(ii). $\forall s \left((T(\dot{T}s) \leftrightarrow T(s^\circ)) \wedge (F(\dot{T}s) \leftrightarrow F(s^\circ)) \right)$;

(iii). $\forall s \left((T(\dot{F}s) \leftrightarrow F(s^\circ)) \wedge (F(\dot{F}s) \leftrightarrow T(s^\circ)) \right)$;

(iv). $\forall x \left(\mathrm{Sent}_{\mathsf{KF}}(x) \rightarrow (T(\dot{\neg}x) \leftrightarrow F(x)) \wedge (F(\dot{\neg}x) \leftrightarrow T(x)) \right)$;
(v). $\forall x \forall y \left(\mathrm{Sent}_{\mathsf{KF}}(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow T(x) \wedge T(y)) \wedge (F(x \dot{\wedge} y) \leftrightarrow F(x) \vee F(y)) \right)$;
(vi). $\forall v \forall x \left(\mathrm{Sent}_{\mathsf{KF}}(\dot{\forall} v x) \rightarrow (T(\dot{\forall} v x) \leftrightarrow \forall t\, T(x[t/v])) \wedge (F(\dot{\forall} v x) \leftrightarrow \exists t\, F(x[t/v])) \right)$.

Remark 3 KF is intimately related to a logical *development of non-extensional concepts* (classification, operation) and *semantical investigations*; see Cantini (1996) for the connections with Aczel's Frege structures, explicit mathematics. Furthermore, KF has an interesting model theory (rich lattice theoretical results); it is also related to the standard fixed point theory $\widehat{\mathsf{ID}_1}$ (see Feferman 1982); in turn, this leads towards the foundations of intuitionistic type theory (see Hancock's conjecture, Feferman 1982, Cantini 1985a).

Halbach (2011) develops a fuller picture. KF can be seen as a generalization of CT, which is a subtheory of KF, or, more naturally, as a generalization of a theory of a positive inductive definition of truth and falsity (see Halbach 2011, § 8.7).

12.5.1.1 Technical Results

For the reader's sake, let me add a reminder on significant fragments and their proof theory.

(i). KF^- is KF with induction for T-free formulas;
(ii). KF_c is KF with induction restricted to total predicates, i.e. such that $\forall x T(a(x) \vee T(\dot{\neg}a(x)))$;
(iii). KF_p is KF with induction restricted to internal predicates;
(iv). CONS is the statement that no sentence in the language $\mathcal{L}_{\mathsf{KF}}$ is true and false.

KF and its variants can be compared with standard subsystems of second order arithmetic as presented e.g. in Simpson's monograph (Simpson, 1999). If $\mathsf{S}_1, \mathsf{S}_2$ are two elementary theories, let $\mathsf{S}_1 \equiv \mathsf{S}_2$ stands for "$\mathsf{S}_1, \mathsf{S}_2$ have the same arithmetical consequences." Then, if $(\Pi^0_1 - \mathrm{CA})_{<\lambda}$ is (an axiomatic rendering of) ramified analysis up to any level $< \lambda$, it is known:

Theorem 2

(i) $\mathsf{PA} \equiv \mathsf{KF}^- + \mathsf{CONS} \equiv \mathsf{KF}_c$;
(ii) KF^- *has nonelementary speed-up over* PA *(see corollary 5.13 of Fischer (2014));*
(iii) $(\Pi^0_1 - CA)_{<\omega^\omega} \equiv \mathsf{KF}_p + \mathsf{CONS}$;
(iv) $(\Pi^0_1 - CA)_{<\varepsilon_0} \equiv \mathsf{KF} + \mathsf{CONS}$.

Let us extend KF *by a suitable substitution rule* of the form

$$\frac{\varphi(P)}{\varphi(\hat{x}\psi(x))}$$

where φ is a formula of $\mathcal{L}_{\mathsf{PA}}$ with an additional predicate symbol P, ψ is arbitrary. Then the resulting system yields the *schematic reflective closure* $\mathrm{Ref}^{*}_{\mathsf{PA}(P)}$, another way to characterize predicativity in the sense of Feferman and Schütte:

Theorem 3

$$Ref^{*}_{\mathsf{PA}(P)} \equiv (\Pi^0_1 - CA)_{<\Gamma_0}.$$

Informally, the rule allows us to make inferences from schemata accepted in the original arithmetical language to schemata of the language with self-referential truth.

12.5.2 Beyond Kripke-Feferman

There is a fairly unsatisfactory feature of KF, which naturally leads to the consideration of consider of a *non-classical* version of KF. Roughly, *internal* truth and *outer* truth diverge in KF; this means that KF can prove a sentence A, and yet KF is unable to prove $T(\ulcorner A \urcorner)$.[17] In other words, KF does not have full disquotation rules T-intro and T-rules.

This fact has motivated the introduction of a *non-classical subsystem* PKF, based on a generalization of partial logic, the so-called BDM-logic (Basic De Morgan logic), which is studied by Halbach and Horsten (2006), Halbach (2011), Horsten (2011), Fischer et al. (2017)). PKF is closed under the rules

- T-introduction $A / T(\ulcorner A \urcorner)$;
- T-elimination $T(\ulcorner A \urcorner)/A$;

hence A and $T(\ulcorner A \urcorner)$ are interderivable.

It has been proved that PKF is proof-theoretically weaker than KF and indeed comparable with KF_p wih respect to its ordinal theoretic content.

To sum up, a dilemma arises. The first horn is that you consistently rely on a notion of *full disquotational truth*, whose naturalness is a good ground for justificatory work. But then the inferential patterns of material implication are not accessible because of the T–rules, which makes mathematical argumentation more cumbersome and restricts its power, whereas mathematically reasoning in classical logic is natural. On the second horn, you can recover standard logic, and state the scientific laws of truth as in KF. But these laws are not assertible in the system, which has to give up a justificatory work for the foundations (for a deepening of this point, see Fischer et al. 2019).

[17] Typically this happens for the Liar sentence L.

12.5.2.1 Autonomous Iterations and Metapredicativity

Type-free truth can be naturally iterated following the simple idea that, when we accept a system S, we are entitled to accept S as true. In other words, we are led to single out the operation

$$S \mapsto S^* \tag{12.5.1}$$

where S^* is S with the axioms on a new predicate T expressing the truth predicate for the language of S and satisfying KF-axioms.

The procedure is then iterated transfinitely many steps along *autonomously obtained primitive recursive well-orderings* prwo $\prec$: roughly you accept $\prec$ of type α only if in a system defined *at a lower stage* you obtain transfinite induction TI along $\prec$. This gives rise to the limit system $Aut(\mathsf{KF})$.

Beyond standard predicativity, one gets ordinals higher up, greater that Γ_0, using the so-called Veblen's ternary function hierarchy $\lambda\alpha\beta\gamma.\varphi\alpha\beta\gamma$ and the function $\lambda\alpha.\Gamma_\alpha$ enumerating the Γ-numbers (see Strahm 2000). Fujimoto's theorems in Fujimoto (2011) provide a proof-theoretic analysis of transfinite iterations and autonomous progressions of various arithmetical truth theories, including $Aut(\mathsf{KF})$, positive uniform T-biconditionals, Kripke-Feferman truth based on a weak Kleene scheme, determinate truth and Feferman logic. A precise characterization in terms of ordinal invariants and all these theories yields the ordinal $\varphi 200$. It is to be noticed that the theory VF (see Cantini 1990) of self-applicable truth based upon the supervaluation schema climbs much higher up, and the ordinal required for $Aut(\mathsf{VF})$ requires the more powerful ordinal function Θ and is known as $\Theta\Omega_{\Omega_1}0$.

The general philosophical meaning of these investigation on self-referential truth predicates is that along this path, one comes to explore systems that go *beyond the boundaries of classical predicativity*. This opens up the route to a fresh analysis of predicativity, also in connection with apparently separate lines of thought, e.g. constructive type theory (for a recent contribution, see Crosilla 2017).

Conceptually, the issue is: how stable is the upper limit of predicativity? At present, there are several converging results providing formal and informal arguments for overcoming the bounds of Feferman-Schütte.

Remark 4 KF and VF have different strength proof-theoretically above PA. This contrasts with a result by Fujimoto (2018) that if one respectively extends NBG with truth axioms à la KF and à la VF, one obtains two theories of the same strength!

12.5.2.2 Completing the Picture

Along the pattern of KF, one can pursue other interesting routes.

(i) Many philosophers think that the minimal fixed point model of Kripke's theory is the most natural: a picture of *grounded* truth. But KF is the theory of *all fixed point models*. Hence one can try to add axioms excluding *non-minimal*

fixed point models and this leads to theories of truth with some minimality assumption (see Burgess 2014, Cantini 1989).

(ii) The theory KF naturally generalizes to the so-called applicative systems, where the ground universe is a combinatory algebra, i.e. a theory of untyped operations, (see Aczel's original contribution (Aczel, 1980) about Frege structures and the previous work by Feferman on explicit mathematics). In recent times (Cantini and Crosilla 2010) semantical theories have been used as an environment to develop an interpretation of constructive set theory CZF, which simplifies Aczel's type-theoretic interpretation and fragments thereof. Thus the roots of the notion of sets lie in the underlying theory of truth and predication, together with a basic structure, where rules (in a generalized computational sense) live. One smoothly moves from truth to sets, constructively. Again, semantical theories give rise to natural predication theories and hence to the logical notion of set à la Frege-Russell.

(iii) One can compare alternative evaluation schemata, as mentioned already (e.g. van Fraassen and the system VF, Feferman's determinate truth Cantini 2017) and try to assess their different use for the philosophy of mathematics.

(iv) A more challenging task can be imagined:

(*) design semantical theories that go beyond a nominalistic predicativistic definitionist view, moving from predicative systems to impredicative systems.

To this aim we have introduced elsewhere (Cantini, 2020) an abstract, very strong theory of truth built up according to Quinean ideas and typical ambiguity; the theory is consistent, if Quine's *New Foundations* is; as to the strength, it goes far beyond the systems investigated e.g. in Schindler (2018).

Acknowledgments I would like to thank the anonymous referees for helpful comments and the editors of this volume, Stefano Boscolo, Gianluigi Oliveri and Claudio Ternullo for their determination in bringing the project of this volume to completion.

References

Aczel, P. 1980. Frege structures and the notions of proposition, truth and set. In *The Kleene Symposium*, ed. J. Barwise, H. Keisler, and K. Kunen, 31–59. Amsterdam: North Holland.

Aczel, P., and M. Rathjen. 2001. Notes on constructive set theory. Technical Report 40, Institut Mittag-Leffler, Djursholm.

Burgess, J. 2014. Friedman and the axiomatization of Kripke's theory of truth. In *Foundational adventures: Essays in honor of Harvey M. Friedman*, ed. N. Tennant, 125–128. London: College Publications.

Caldon, P., and A. Ignjatovic. 2005. On mathematical instrumentalism. *Journal of Symbolic Logic* 50: 778–794.

Cantini, A. 1985a. A note on a predicatively reducible theory of iterated elementary induction. *Bollettino dell'Unione Matematica Italiana* 6(2): 1–17.

Cantini, A. 1985b. Una nota sulla concezione semi–intuizionistica della matematica. *Rivista di Filosofia* 69: 465–486.
Cantini, A. 1989. Notes on formal theories of truth. *Zeitschrift f. Math. Logik u. Grundlagen der Mathematik* 35: 97–130.
Cantini, A. 1990. A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic* 55: 244–259.
Cantini, A. 1996. *Logical Frameworks for Truth and Abstraction*. Amsterdam: North Holland.
Cantini, A. 2016. About truth and types. In *Advances in proof theory*, volume 28 of *Progress in computer science and applied logic*, ed. R. Kahle, T. Strahm, and T. Studer, 287–314. Cham: Birkhäuser–Springer.
Cantini, A. 2017. Feferman and the truth. In *Feferman on foundations*, volume 17 of *Outstanding contributions to logic*, ed. G. Jäger and W. Sieg, 31–64. Cham: Springer.
Cantini, A. 2020. A fixed point theory over stratified truth. *Mathematical Logic Quarterly*. to appear, 17 pages.
Cantini, A., and L. Crosilla. 2010. Elementary constructive operational set theory. In *Ways of proof theory*, ed. R. Schindler, 199–240. Frankfurt: Ontos Verlag.
Cellucci, C. 2017. *Rethinking knowledge: The heuristic view*. Cham: Springer.
Cieśliński, C., M. Łełyk, and B. Wcisło. 2017. Models of PT^- with internal induction for total formulae. *Review of Symbolic Logic* 10: 187–202.
Crosilla, L. 2017. Predicativity and Feferman. In *Feferman on foundations*, volume 17 of *Outstanding contributions to logic*, ed. G. Jäger, and W. Sieg, 423–447. Cham: Springer.
Dummett, M. 2004. *Truth and the past*. New York: Columbia U.P.
Ehrenfeucht, A., and J. Mycielski. 2020. Abbreviating proofs by adding new axioms. *Bulletin of the American Mathematical Society* 77: 366–367.
Enayat, A., M. Lełyk, and B. Wcisło. 2020. Truth and feasible reducibility. *Journal of Symbolic Logic* 85: 367–421.
Enayat, A., and A. Visser. 2015. New constructions of satisfaction classes. In *Unifying the philosophy of truth*, Number 36 in Log. Epistemol. Unity Sci., ed. T. Achourioti, H. Galinon, J.M. Martinez, and K. Fujimoto, 321–35. Cham: Springer.
Feferman, S. 1979. Constructive theories of functions and classes. In *Logic colloquium '78*, ed. M. Boffa and D. van Dalen, 159–224. Amsterdam: North Holland.
Feferman, S. 1982. Iterated inductive fixed-point theories: Application to Hancock's conjecture. In *Patras Logic Symposion '80*, ed. G.E.A. Metakides, 171–196. Amsterdam: North Holland.
Feferman, S. 1998a. *In the light of logic*. Oxford: Oxford University Press.
Feferman, S. 1998b. Weyl vindicated: Das Kontinuum 70 years later. In *In the light of logic*, 249–283. Oxford: Oxford University Press.
Feferman, S. 1998c. Working foundations '91. See Feferman (1998a), 105–124.
Feferman, S. 2008. Axioms for determinateness and truth. *Review of Symbolic Logic* 1: 204–217.
Fischer, M. 2014. Truth and speed-up. *Review of Symbolic Logic* 7: 319–340.
Fischer, M., L. Horsten, and C. Nicolai. 2017. Iterated reflection over full disquotational truth. *Journal of Logic and Computation* 27: 2613–2651.
Fischer, M., L. Horsten, and C. Nicolai. 2019. Hypatia's silence. *Nous* 20: 1–24.
Fujimoto, K. 2011. Autonomous progressions and transfinite iteration of self-applicable truth. *Journal of Symbolic Logic* 76: 914–945.
Fujimoto, K. 2018. Truths, inductive definitions, and Kripke–Platek systems over set theory. *Journal of Symbolic Logic* 83: 868–898.
Fujimoto, K. 2019. Predicativism about classes. *Journal of Philosophy* 116: 206–229.
Girard, J. 1987. *Proof theory and logical complexity*. Naples: Bibliopolis.
Glass, T. 1996. On power set in explicit mathematics. *Journal of Symbolic Logic* 61(2): 468–489.
Gödel, K. 1986. Über die länge von beweisen. In *Collected works*, Vol. I, ed. S. Feferman and et al., 123–153. Oxford: Oxford University Press.
Halbach, V. 2011. *Axiomatic theories of truth*, 1st ed. Cambridge: Cambridge University Press.
Halbach, V., and H. Horsten. 2006. Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic* 71(2): 677–712.

Hilbert, D. 2013. *David Hilbert's lectures on the foundations of arithmetic and logic, 1917–1933*. Cham: Springer.

Horsten, L. 2011. *The Tarskian turn. Deflationism and axiomatic truth*. Cambridge: MIT Press.

Horsten, L., and G. Leigh. 2017. Truth is simple. *Mind* 126: 195–232.

Jäger, G. 1997. Power types in explicit mathematics? *Journal of Symbolic Logic* 62: 1142–1146.

Jäger, G. 2005. Metapredicative and explicit Mahlo: A proof-theoretic perspective. In *Logic colloquium 2000*, Lecture notes in logic 19, ed. R. Cori, A. Razborov, S. Todorcevic, and C. Wood, 272–293. Urbana: Association for Symbolic Logic.

Jäger, G., R. Kahle, A. Setzer, and T. Strahm. 1999. The proof-theoretic analysis of transfinitely iterated fixed point theories. *Journal of Symbolic Logic* 64: 53–67.

Kaye, R. 2005. *Models of peano arithmetic*. Lecture notes in logic 19. Urbana: Association for Symbolic Logic.

Kotlarski, H., and A. Lachlan. 1981. Construction of satisfaction classes for nonstandard models. *Canadian Mathematical Bulletin* 24: 283–93.

Kreisel, G. 1970. Principles of proof and ordinals implicit in given concepts. In *Intuitionism and Proof Theory*, ed. A. Kino, J. Myhill, and R.E. Vesley, 489–516. Amsterdam: North Holland.

Kripke, S. 1975. Outline of a theory of truth. *Journal of Philosophy* 72: 670–716.

Lachlan, A. 1981. Full satisfaction classes and recursive saturation. *Canadian Mathematical Bulletin* 24: 295–297.

Leigh, G.E. 2015. Conservativity for theories of compositional truth via cut elimination. *The Journal of Symbolic Logic* 80: 845–865.

Łełyk, M., and B. Wcisło. 2017a. Models of weak theories of truth. *Archive for Mathematical Logic* 56: 453–474.

Łełyk, M., and B. Wcisło. 2017b. Notes on bounded induction for the compositional truth predicate. *Review of Symbolic Logic* 10: 455–480.

Mancosu, P. 1998. *From Brouwer to Hilbert the debate on the foundations of mathematics in the 1920s*. Oxford/New York: Oxford University Press.

Parsons, C. 1971. Ontology and mathematics. *The Philosophical Review* 80: 151–176.

Parsons, C. 1974. Sets and classes. *Noûs* 8: 1–12.

Parsons, C. 2002. Realism and the debate on impredicativity. 1917–1944. In *Reflections on the foundations of mathematics*, Lecture notes in logic 15, ed. W. Sieg, R. Sommer, and C. Talcott, 372–389. Natick: ASL and A.K.Peters.

Rathjen, M. 2016. Indefinitess in semi-intuitionistic set theories: on a conjecture of Feferman. *Journal of Symbolic Logic* 81: 742–754.

Russell, B. 1906. On somes difficulties in the theory of transfinite numbers and order type. *The Proceedings of the London Mathematical Society* 4: 29–53.

Russell, B. 1919. *Introduction to mathematical philosophy*. London: Allen and Unwin.

Schindler, T. 2018. Some notes and comprehension. *The Journal of Philosophical Logic* 47: 449–479.

Simpson, S. 1999. *Subsystem of second order arithmetic*. Berlin/Heidelberg: Allen and Unwin.

Strahm, T. 2000. Autonomous fixed point progressions and fixed point transfinite recursion. In *Logic colloquium '98*, ed. S. Buss, P. Hájek, and P. Pudlák, 449–464. Association for Symbolic Logic.

Tait, W. 1983. Against intuitionism: constructive mathematics is part of classical mathematics. *The Journal of Philosophical Logic* 12: 449–479.

Tait, W. 1986. Truth and proof: The platonism of mathematics. *Synthese* 69: 341–370.

Tarski, A. 1944. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4: 341–376.

Tarski, A. 1956. The concept of truth in formalized language. In *Logic, Semantics, Metamathematics: Papers From 1923 to 1938*, ed. J. Woodger, 152–278. Oxford: Clarendon Press.

Weyl, H. 1910. Über die Definitionen der mathematischen Grundbegriffe. In *Gesammelte Abhandlungen* vol.1, Mathematisch-naturwissenschaftliche Blätter 7:93–95/109–113, 298–304.

Chapter 13
On Lakatos's Decomposition of the Notion of Proof

Enrico Moriconi

Abstract Lakatos's masterpiece, *Proofs and Refutations*, can be seen as a renewal of the classical debate between analytical and axiomatic (or synthetic) procedures. In fact, it is framed within a broad context where the refusal of the latter kind of procedures is linked to a strong criticism of the formalistic school of mathematical philosophy.

In this paper I will try to appropriately settle the relationship between Popper's falsificationism and Lakatos's fallibilism, and I will give a special emphasis to the final part of *Proofs and Refutations*, where the last notion to be submitted to stretching becomes that same notion, when Lakatos starts "to stretch the stretching", and the skeptic component of his attitude becomes the source of some of his most brilliant insights.

Keywords Formal/informal proofs · Falsificationism · Fallibilism

13.1 Introduction

Lakatos's masterpiece, *Proofs and Refutations*,[1] constitutes a wonderful exploration of how mathematics evolves. The central theme of Lakatos's dissertation[2] is a criticism of the concept of *formal proof*, which is an argument executed according

[1] See Lakatos (1976). From now on, we will use $\mathcal{P}\&\mathcal{R}$ as a shorthand for the volume *Proofs and Refutations*.

[2] As is well known, $\mathcal{P}\&\mathcal{R}$ is written as a classroom dialogue between a teacher (Lakatos?) and students, simply recorded with a Greek letter. Students, usually, raise new questions and problems providing also the (often) temporary answers and solutions, whereas the teacher just takes stock of the debate highlighting points which are worthy of further discussion.

E. Moriconi (✉)
Department of Civilisations and Forms of Knowledge, via P. Paoli 15, Pisa, Italia
e-mail: enrico.moriconi@unipi.it

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_13

to the rules of a precisely specified mechanism. His aim is to give some real sense to the claim that regimenting proofs in order to clarify their assumptions and the procedural rules involved –the process which formalization brings to the fore– is just *one* phase in the complex process that leads to the growth of mathematical knowledge.

After having detailed the few (quasi-)technical terms one has to get acquainted with in order to understand Lakatos's *fallibilist* proposal, I will focus on the modifications that, according to a largely shared opinion, Lakatos produced on Popper's critical *falsificationism* with regard to two core aspects:

1. Lakatos extends falsificationism also to mathematics (to which Popper himself did not venture to apply his ideas), proposing to consider also the latter as quasi-empirical: mathematical theorems are not irrefutably true statements, but conjectures: one cannot know that a theorem will not be refuted.
2. According to Lakatos, refutation does not entail immediate rejection, as it was the case in Popper's Darwinistic account. Lakatos deploys instead a battery of strategic moves in order to cope with cropping out counterexamples.

The previous characterizations of Lakatos's views contain significant portions of truth, of course, but I don't believe that they offer a very balanced account of things, and I think that Popper's logical and epistemological papers of the late 1940s can help us in better setting the relationship between Popper's falsificationism and Lakatos's fallibilism. I will ground my ideas in positions held by Popper in various papers of the years from 1946 to 1948, most notably in "Why are the Calculuses of Logic and Arithmetic Applicable to Reality?" (1946), and in "Logic without Assumptions" (1947), written when Popper was teaching the 2-year introductory courses on logic and scientific method which were his core academic duty at the *London School of Economics* for over two decades (1946–1969). For, I think that attending Popper's courses and seminars, together with acquaintance with his logical papers, was a true "training ground" for most of his students, most probably Imre Lakatos included.

In fact it is difficult not to see a close connection between Popper's *playing* with the logical notions and Lakatos's perspective, in which the starting point is got by adapting a somehow devised "proof-sketch" to new problems waiting for a solution. A main point, in this framework, is given by the way in which the evolution of the initial proof of Euler's Conjecture is depicted as something which results from interactions with various kinds of counterexamples which immediately start to crop up, leading to arguments over the *meaning* of terms involved in the definitions as they are put forward, so that various characterizations of the notions of polyhedron, polygon, edge, area, vertex, …, are provided. For instance, when student Alpha proposes the *hollow cube* as a counterexample to the proof, he causes a discussion in the classroom which produces a change in the definition of "polyhedron": from

> A polyhedron is a solid whose surface consists of polygonal faces. (p. 14)

to

> A polyhedron is a surface consisting of a system of polygons. (ib.)

A few pages later, we see that the discussion of Kepler's star-polyhedron (the so-called *urchin*), leads the classroom to question the notion of "face" (pp.17–18), and the examination of the *cylinder* (p. 22) leads the classroom to wonder about the notion of "edge". The theoretical torsion which the various notions undergo is shaped on the first physical, so to speak, step of the proof (or thought-experiment), due to Cauchy, of the conjecture. This consists in removing one face of the polyhedron and then *stretch* it flat onto a blackboard. From here on, *stretching* becomes a usual practice of the text, for instance, when a concept definition is modified so as to exclude an unwanted counterexample,[3] or when one reinterprets a counterexample so that it no longer violates Euler's Conjecture.[4]

In this paper, however, I will not follow the various phases of the discussion held in the classroom. The focus of my paper will be instead on the last part of $\mathcal{P}\&\mathcal{R}$, where the last notion to be submitted to stretching becomes this very notion, and Lakatos starts "to stretch the stretching".[5] In this way the skeptic component of Lakatos's attitude resurfaces in the claim that no single language can model the growth of knowledge, and that there is no hope that the mechanism of refutational success, i.e. "concept-stretching", could peter out. I will show that this same attitude is the source of some of his most brilliant insights.

13.2 Injecting Truth and Meaning

$\mathcal{P}\&\mathcal{R}$ can be seen as a renewal of the classical debate between analytical and axiomatic (or synthetic) procedures. In fact, it is framed within a broad context where the refusal of the latter kind of procedures is linked to a strong criticism of the traditional *euclidean*, or *axiomatic*, and *formalistic* school of mathematical philosophy. It is not relevant here to fully detail the description of traditional perspectives, so we will limit ourselves to remind just the basic notions that provide the necessary background to settle Lakatos's proposal.

In Euclidean Axiomatics, injection of truth and meaning is given at the *outset* by means of definitions (which state which kind of entities we will dwell upon) and postulates, or axioms (whose truth, guaranteed by intuitive evidence, is propagated to any other assertion).

In Hilbertian Axiomatic Formal Systems both questions are displaced at the very *end* of the construction. As regards the question of truth, it has been, so to speak,

[3] And in this case one speaks of the method of *Monster barring*.

[4] And in this case one speaks of *Monster adjusting.*

[5] At p. 102, student Gamma answers Kappa back by saying: "You stretch the concept of concept-stretching!"

ignored and replaced by the requirement of consistency. As regards the question of meaning, it has been postponed to the formulation of the axioms: for instance, what it means to be an Euclidean point (or line, or plane, ...) is something which is (implicitly) fixed by the axioms.

Lakatos intends to occupy a *land* between Euclidean and Hilbertian perspectives: truth and meaning are injected in the course of the inquiry, without any hope to reach an ultimate point. Both notions have a tentative, conjectural nature.

- As regards *truth*: during the dialogue, after pupil Gamma has reminded that the very term "polyedron" has been stretched to the point that it does not figure in the theorem anymore, pupil Kappa adds

  > because of concept-stretching, refutability means refutation. So you slide on to the infinite slope, refuting *each* theorem and replacing it by a more "rigorous" one, by one whose falsehood has not been "exposed" yet! But *you never get out of falsehood* (p. 100).

- As regards *meaning*: if one tries to stop the previous "infinite slope" by exploiting monster-barring definitions, and so try to keep away counterexamples generated by concept-stretching, pupil Kappa warns that

  > you will slide on to another infinite slope: you will be forced to admit of each "particular linguistic form" of your true theorem that it was not precise enough, and you will be forced to incorporate in it more and more "rigorous" definitions couched in terms whose vagueness has not been exposed yet! But *you never get out of vagueness* (p. 100).

In opposition to both kinds of what he calls *traditional perspectives*, Lakatos proposed to assimilate mathematical theories to general scientific theories, turning them into *quasi-empirical theories*. In the *Introduction* to $\mathcal{P}\&\mathcal{R}$ Lakatos argues that he will challenge mathematical formalism elaborating

> the point that informal, quasi-empirical, mathematics does not grow through a monotonous increase in the number of indubitably established theorems but through the incessant improvement of guesses by speculation and criticism, by the logic of proofs and refutations (p. 5).

13.2.1 The Logic of $\mathcal{P}\&\mathcal{R}$

There are a few (quasi-)technical terms one has to get acquainted with in order to understand Lakatos's proposal.

1. Central for Lakatos's philosophy of mathematics is his characterization of the concept of mathematical **proof**, which occurs near the beginning of the text:

   > Teacher: I propose to retain the time-honoured technical term 'proof' for a thought-experiment –or 'quasi-experiment'– which suggests a **decomposition** of the original conjecture into subconjectures or lemmas, thus embedding it in a possibly quite distant body of knowledge (p. 9).

2. After "proof", we remind the notion of **proof analysis**, which means the production of what we might normally call the "proof": the list of "lemmas" into which the proof (thought-experiment) is decomposed. We are doing proof analysis when we study the precise conditions under which the moves taken in the proof can be made, or are correct.
3. An important role is played by the notions of **local counterexample** and **global counterexample**. Global counterexamples show that some universal statement is false, but in a way that does not require any reference to the proof of that statement. It is a sort of counterexample which isn't at odds with any step produced by the proof analysis. A local counterexample, by contrast, is a property pertaining not to a statement but to a proof of a given statement. Thus the definition of a local counterexample refers both to a statement and to a proof of it, regarded as a sequence of other statements.
4. The goal of the development of a proof, like that of Euler's formula, is a rigorous *theorem*, which Lakatos calls the **principle of retransmission of falsity**, meaning that all *global* counterexamples must become (also) *local*. Falsity must be retransmitted from the naive conjecture (decomposed by the proof) to the lemmata (provided by the proof analysis). That is, any counterexample to the theorem should be a counterexample to some step in the proof-analysis of the theorem:

   > Lambda: A proof-analysis is 'rigorous' or 'valid' and the corresponding mathematical theorem true if, and only if, there is no 'third-type' counterexample to it (p. 47).

   (We remind that the third-type counterexamples are those that are global – they refute the theorem at hand – but not local – they do not falsify any step of the proof.)
5. The last notion to consider is the **principle of the retransmission of truth**, a notion which pertains to the case of counterexamples which are both local and global. The hollow cube, for instance, which is a cube with a cube shaped hole in it, is a counterexample which is both global (since $V - E + F = 16 - 24 + 12 = 4$), and local (since it cannot be stretched flat on the blackboard having had a face removed). To treat this type of counterexamples the faulty lemma is made up a condition of the original conjecture, restricting in this way its range of applicability. The proof is left unchanged, and just like with the question of the convexity, in this case too we have no assurance that even some polyhedron which does not satisfy the lemma (become part of original conjecture) is still an Eulerian polyhedron.

It is tempting to see the last two notions as very kin to, respectively, the *soundness* and the *semantic completeness* of a (formal) theory, the properties which impede the overgeneration and undergeneration of mathematical truths.

However, because of the peculiarities of Lakatos's perspective, this is a temptation we must resist. The evolution of the initial "proof" sketch, or thought-experiment, results from interactions with various kinds of counterexamples. At each stage the proposed counterexamples are examined evaluating the reasons for

the possible inadequacy of the proof, where such an examination may provide hints as to how both the steps and the notions involved in the proof can be modified.

The program that in this way comes to light is that of an *open framework*: not necessarily the described procedure produces a convergent sequence of proofs flowing into a definite, ultimate proof, least of all in a definite proof of the original conjecture. And the possibility to abandon the original conjecture cannot be excluded.

$\mathcal{P}\&\mathcal{R}$ takes into account various ways of coping with counterexamples. It would be however a serious mistake to search for *the* correct method. The correct perspective is precisely given by the interplay of different methods to face different kinds of counterexamples; i.e., the interplay between generation and evaluation of counterexamples. What is worth to stress is that both generation and evaluation are driven by the proof. As student Beta admits, the logic of conjectures and refutations has no starting point (naive conjectures are preceded by many "pre-naive" conjectures and refutations), but the logic of proofs and refutations has: it starts with the first naive conjecture to be followed by a proof, intended as a thought-experiment (p. 74).

13.3 Popper Comes into Play

As we said in the Introduction, according to a largely shared opinion, Lakatos was able to shape his own research project by *distancing Popper*, that is, by modifying Popper's critical falsificationism with regard to two core aspects: first, he extended falsificationism also to mathematics, and, second, replaced Popper's strict notion of refutation with a much more elaborate battery of strategic moves in order to cope with cropping out counterexamples.

As I said, the relationship between Popper's falsificationism and Lakatos's fallibilism is a question which deserves a better consideration. As regards the extension of the falsificationist perspective also to mathematics, in fact, interesting and relevant suggestions are already present in Popper (1946). Facing the central question: "Why are the logical calculi – which may contain arithmetic – applicable to reality?", he notes that certain calculi – for example, the arithmetic of natural numbers, or that of real numbers – are helpful in describing certain kinds of facts, but not other kinds; and he adds:

> In so far as a calculus is applied to reality, it loses the character of a *logical* calculus and becomes a descriptive theory *which may be empirically refutable*; and in so far it is treated as irrefutable, i.e., as a system of *logically true* formulae, rather than a descriptive scientific theory, is not applied to reality (p. 54).

These substantial hints towards the possibility to frame also mathematics within the falsificationist perspective are then deepened in a significant way for our argument when Popper notes that a proposition such as "2 + 2 = 4", if applied to apples, is taken to be irrefutable and logically true, but it does not describe any fact involving apples – any more than the statement "All apples are apples" does. Like

this latter statement, it is a logical truism; and the only difference is that it is based, instead of on the definitions of the signs "All" and "are", on certain definitions of the signs "2", "4", "+", and "=". (Definitions which may be either explicit or implicit.) We might say in this case that the application is not *real* but only *apparent*; that we do not describe here reality, but only assert that a certain way of describing reality is equivalent to another way.

However, there is a second sense in which one could interpret a sentence such as "2 + 2 = 4": it may be taken to mean that, if somebody has put two apples in a certain basket, and then again two, and has not taken any apples out of the basket, there will be four in it. In this interpretation "2 + 2 = 4" helps to calculate, i.e. to describe certain *physical facts*, and the symbol "+" stands for a *physical manipulation* – for physically adding certain things to other things. But in this interpretation the statement "2 + 2 = 4" becomes a physical theory, rather than a logical one; and as a consequence, we cannot be sure whether it remains universally true.

Very easily Popper provides examples of models in which "2 + 2 = 4" is not applicable (a couple of rabbits of different sexes, or a few drops of water, . . .), and to criticism based on the convinction that the equation "2 + 2 = 4" only applies to objects to which nothing happens, he replies that then that equation does not hold for "reality" (for in "reality" something happens all the time), but only for an abstract world of distinct objects in which nothing happens.

And Popper concludes maintaining that

> whenever we are doubtful whether or not our statements deal with the real world, we can decide it by asking ourselves whether or not we are ready to accept an empirical refutation. If we are determined, on principle, to defend our statements in the face of refutations [. . .], then we are not speaking about reality. Only if we are ready to accept refutations do we speak about reality (Popper, 1946, p. 56).

Pertinent to the previous remarks is also the following passage, taken from unpublished notes of 1954–1955, quoted by Bar-Am in Bar-Am (2009), which reveals much about the character of the course in logic and scientific method Popper taught at LSE from 1946 to 1969:

> The idea that science "proves" is wrong. The word "proves" is being misunderstood. In the sense of "prove" discussed above science has "proved" very little. Look at the changes in science in the last 2,000 years. If on important points science can change its teaching so much in the course of time, the proof, if it occurs at all in science must be comparatively rare . . . it marked a kind of false idea in science, an idea of science in which science cannot change, only grow.

Lastly, I think it is appropriate also to mention the sort of *manifesto* which we find in the very first lines of Popper (1947), and which is as good for him as for what Lakatos was going to do a few years later. Confronted with the *problem of giving a satisfactory definition of "valid inference"*, Popper says that

> Our method will be as follows: after having introduced, in section (1), a few auxiliary technical terms, we shall propose a definition, criticize it, and replace it by a better one, and repeat this procedure. (p.251)

The affinity between this position and Lakatos's stance is impressive.

13.4 Back to Lakatos

I think that attending Popper's courses and seminars was a true "training ground" for most of his students, Imre Lakatos included.[6]

It is in fact difficult not to see a deep link between Popper's *playing* with the logical notions[7] and Lakatos's perspective, in which, I remind, the starting point is got by adapting a somehow devised "proof-sketch" to new problems waiting for a solution.

13.4.1 Concept-Stretching

On p. 93 of Lakatos (1976) pupil Pi notes that it is impossible to keep refutations and proofs on the one hand and, on the other hand, changes in the conceptual, taxonomical, linguistic framework. Facing a counterexample, one can choose to disregard it because it has to do with notions not belonging to his language $\mathcal{L}_1$, or to accept the counterexample passing by concept-stretching to a new language $\mathcal{L}_2$, ... Commenting on this point, pupil Gamma foresees the possibility of *inconsistent* languages:

> we may have two statements that are consistent in $\mathcal{L}_1$, but we switch to $\mathcal{L}_2$ in which they are inconsistent. Or, we may have two statements that are inconsistent in $\mathcal{L}_1$, but we switch in $\mathcal{L}_2$ in which they are consistent. [...] The growth of knowledge cannot be modeled in any *given* language [my emphasis].

The skeptic component of Lakatos's attitude surfaces in the claim that no single language can model the growth of knowledge, and that there is no hope that the mechanism of refutational success, i.e. the mechanism of "concept-stretching", could peter out. This same attitude is the source of some of his most brilliant insights.

This is also the point where Lakatos goes beyond Popper's bounds of reasoning. In a similar vein, in fact, and of course with reference to the consequences of Gödel's results of 1931,[8] in Popper (1946) Popper maintains that, although it is possible, for *any* given valid intuitive inference, to construct some language permitting the formalisation of that inference, it is not possible to construct *one* language or calculus allowing us to formalise *all* valid intuitive inferences, provided of course that we do not exploit procedures of an entirely different character. That is to say, procedures which, as the so-called ω-rule, allow inferences which can be drawn

[6] I think important to remember that Lakatos's paper was first read in Karl Popper's seminar in London in March 1959.

[7] See Binder and Piecha (2017) for an exhaustive and very clear exposition of Popper's investigations. Other details are available in Moriconi (2019).

[8] And to the way they were acknowledged by Tarski in (2002).

from an *infinite* class of premises. These remarks by Popper, however, point to the openness of the research, without making concessions either to skepticism or to meaning variance. Also Lakatos's remarks are strictly connected to the results of incompleteness, and to the consequent failing of categoricity of first order theories. In *What does a mathematical proof prove?*, which is more or less contemporary to Lakatos (1976),[9] in fact, he stresses that as a consequence of those proofs, we cannot fully control formal proofs, meaning that we cannot avoid that they prove much more than we want them to prove, without being possible to impede that the proof of an arithmetical theorem turns at the same time into the proof of some theorem in other absolutely *unintended* structures. To put it from the point of view of the system of axioms, this entails that we cannot avoid that the axioms in the most important mathematical theories implicitly define not just one, but quite a family of structures, with the concrete possibility that there exist statements which are true in one structure but false in the other. "Truth value" and "meaning" continue to proceed hand in hand, without the possibility to disentangle the structure we intended from the multiplicity implicitly characterized by our axiomatic framework.

As we already reminded, pupil Kappa warns about the two infinite slopes one risks to slide onto: on the one hand, the slope induced by the attempt to replace *each* refuted theorem by a more "true" one; on the other hand, the slope generated by the attempt to reach the most *precise* linguistic form of the theorem. There is, however, no hope to reach a stop: "*you never get out of falsity*", and "*you never get out of vagueness*". And he seals his argument by stressing that

> For any proposition there is always some sufficiently narrow interpretation of its terms, such that it turns out true, and some sufficiently wide interpretation such that it turns out false. Which interpretation is intended and which unintended depends of course on our intentions. [...] Concept-stretching will refute *any* statement, and will leave no true statement whatsoever (p. 99) [My emphasis].

13.4.2 I Shall Stretch "Stretching"

The last notion to be submitted to "stretching" is therefore this very same notion, and it is intruiguing to compare the previous quoted assertion with the following quotation from Wittgenstein (*Philosophical Investigations*,§ 201):

> [...] *no* course of action could be determined by a rule, because *every* course of action can be made out to accord with a rule.

This is the assertion which famously constitutes the starting point of the argument developed by Kripke in his *Wittgenstein on Rules and Private Language* of 1982,[10] and it promises to be very interesting to consider this part of the dialogue keeping in mind the themes of Kripke's book. Clearly, Lakatos does not show interest in

[9] See Lakatos (1978).

[10] Kripke (1982).

questions concerning the topic of "private language": the line of his argument which is here pertinent, in fact, is that which hinges on the distinction between intended and unintended interpretations.[11]

And the closeness with issues that will be employed by Kripke – especially with the mathematical example Kripke provides to support his interpretation of the rule-following paradox– is particularly noteworthy when Lakatos, following the thread of skepticism, arrives to stretch the meaning of the arithmetical (as for instance, addition) and logical (as for instance, the universal quantifier) notions.

13.4.3 Stretching Arithmetical Notions

Answering to pupil Gamma, who hopes to reach a point where the meanings of the terms will be so crystal clear that there will be one single interpretation, as it is the case with $2 + 2 = 4$, pupil Kappa shows that it is possible to stretch also this proposition by stretching the meaning of "addition". To this aim it is envisaged a generalized notion of addition which could be called *addition as package*. The usual addition[12] is recovered from that as the very special case of packing where the weight of the covering material is zero.

Like Kripke's skeptic, also Lakatos's skeptic does not challenge the normal arithmetical operations, but differently from the former, he does not search for evidence that in making calculations what is really intended is "usual addition" and not "addition as package".[13] He does not complain about the lack of some fact of the matter (either physical or mental, like "dispositions" to react in a certain way) that can constitute the state of meaning supporting the usual addition rather than the addition as package. His skepticism is much more kin to the classic one, *empirically* coloured. As Kappa says:

> *But there are no [inelastic, exact] concepts! Why not accept that our ability to specify what we mean is nil, therefore our ability to prove is nil?* [...] If you want mathematics to be meaningful, you must resign of certainty. If you want certainty, get rid of meaning. You cannot have both (p. 102).

From our point of view, beyond looking forward, it is also interesting to look backward: it is in fact immediate to remind (Popper, 1946), where we encountered an analogous variation on the operation of addition, but of course absolutely lacking any skeptic trace.

[11] Few pages before, the character Pi said

> The [Euler's] conjecture was true in its *intended interpretation*, it was only false in an *unintended interpretation* smuggled in by the refutationists. (pp. 84–5).

[12] "[T]he addition in the originally *intended* sense" (p. 102), as Kappa points out [my emphasis].

[13] Assuming that in both cases we question about which rule is understood given a *finite* amount of behaviour.

Popper's speculation on the applicability of logic and arithmetic to reality was, so to speak, the breach through which Lakatos could insert his quasi-empirical proposal *skeptically* coloured.[14]

13.4.4 Stretching Logical Notions

Arguing about the cylinder, and wondering which kind of counterexample to the "Cauchy proof" it is, pupil Gamma claims that the falsity of "there is a diagonal of the circle that does not create a new face" entails the truth of its negation, namely of "all diagonals of the circle create a new face" (p. 44). Lakatos is well aware that the latter sentence has to be treated cautiously since impinging on the basic relation "truth value-meaning": pupil Alpha, in fact, stresses that, being a universal sentence which cannot be instantiated, one could draw the consequence that far from being *true*, it is *meaningless*. The negation of a meaningless sentence, however, cannot be false, but it would result *meaningless* in its turn. As a way out from this impasse, Alpha provides a reformulation the original universal sentence in something like: "for all x, if x is a diagonal, then x cuts the face into two; and there is at least one x that is a diagonal", where the existential clause *hidden* in the lemma is made explicit.[15] This is the move appropriate to make the rules of the Aristotelian square work: being a conjunction, say of the form $\forall x(D(x) \rightarrow C(x))\&\exists x D(x)$, the latter sentence is false, due to the falsity of the second conjunct. Therefore, it follows the truth of its negation, say of $\exists x(D(x)\&\neg(C(x))) \vee \neg\exists x D(x)$, due to the truth of the second disjunct. Confronted with Alpha's manoeuvre, Gamma maintains the meaningfulness of his position, consisting in the acknowledgment of the existence of *vacuously true statements*. And Lakatos highlights how remarkable was also this *modest stretching* underwent by the universal quantifier, consisting in removing the existential import from its meaning, so that it no longer applies only to non-empty classes.[16] And he emphasizes that this was an important event, since it draws attention on the possibility that also logical notions experience

[14] In Popper (1946) we find that Popper wonders why ought we avoid those breaches of the rules of logic that we call "fallacies" if not because we are interested in formulating or deriving statements which are true, that is to say, which are true descriptions of facts?

We undertake the "meta-linguistic" task of detecting the rules of inference of the language we are investigating because we aim at *formalising all* those inferences which we intuitively *know how* to draw; much as we know that it is impossible to build a single calculus able to formalise *all* valid intuitive rules of inference.

[15] And this is good, because the cylinder, from being a global but *not* local counterexample (the third type), in this way becomes a global *and* local counterexample (the second type), abiding by the "Principle of Retransmission of Falsity" (see point 4 of §2.1).

[16] In fact, he notes that

> turning the empty set from a monster into an ordinary *bourgeois* set was an important event –connected not only with the Boolean set-theoretical re-interpretation of Aristotelian

some shifts of meaning. Moreover, and most importantly, this is the point where "stretching" is connected to the process of assessment of the notion of *logical truth*. Stretching the meaning of the universal quantifier, "so that it no longer applied only to non-empty classes", in fact, is a move which is framed within a short description of the steps which brought to the characterization of the notions of logical consequence and logical truth. Lakatos mentions Bolzano and Tarski, of course, and their (successful?) attempt to characterize as logically true a proposition when its truth depends only on the *meaning* of

> those terms whose meaning can be stretched only at the cost of destroying the basic principles of rationality (p. 103).

Lakatos, however, doesn't miss to point to the difficulties encountered in finding a demarcation line between stretchable, or *descriptive*, terms and unstretchable, or *logical* or, to use Popper's terminology, *formative*, terms, and acknowledges to Popper's *inferential* definitions of the logical notions –reference is to Popper (1947)– the merit to have firmly linked those notions to "some basic principles of rational discussion" (p. 104).

In Chap. 2, added by the Editors to the original *Proofs and Refutations*, Lakatos goes back to this subject and draws attention to a new point, different from the question of how to distinguish formative from descriptive signs. It is the fact that in listing and characterising the formative signs one has to introduce a certain amount of artificial limitation and rigidity which is quite foreign to naturally grown languages and to intuitive inferences they allow to build. And he maintains that the formalized concepts of logical consequence and truth, which until now were generally used in the construction of deductive theories, by no means coincide with the everyday concepts, with the everyday "pre-existing" way they are used. In so doing, Lakatos is actually very near to Tarski's perplexities concerning the possibility to *sintactically* catch the essential content of the everyday notions of logical consequence and logical truth,[17] insisting that due attention must be paid also to the translation rules which transfer the intuitive inferences to the formal framework. He remembers that to overcome Tarski's puzzle, which comes to the fore in the final part of his 1936 paper "On the notion of logically following", Popper tried to reverse Tarski's order of priority, taking the notion of "derivability" as primitive, and showing that those signs are *logical*, or *formative*, which can be defined with the help of that primitive concept. But it is clear that to find the correct

---

> logic, but also with the emergence of the concept of vacuous satisfaction in mathematical discussion. (p. 104).

[17] See, for instance, the following Tarskian assertion:

> no matter how we enrich the stock of rules of inference- we shall be able to construct sentences which follow in the everyday sense from the theorems of the deductive theory under consideration, but which cannot be proven in this theory on the basis of the accepted rules. (Tarski (2002), § 1.4.2)

order of priority is just *one* problem, and moreover a problem which Lakatos doesn't appreciate so much.

In a footnote on p. 123, Lakatos observes that an unsatisfactory trait of Popper's treatment of logical form is the unsufficient attention devoted to the important problem of translatory definitions. The limit of Popper's idea is to search for a definition of valid inference depending *only* on the list of formative signs, whereas

> [V]alidity of an intuitive inference depends also on translation of the inference from ordinary (or arithmetical, geometrical, etc.) language into the logical language: it depends on the translation we adopt.

Accordingly, Lakatos reminds that the case of Euler's Conjecture –which was of the form "All *A*'s [polyhedrons] are *B*'s [Eulerian]"– is in fact evidence that assessing logical validity does not hinge *only* on the list of formative signs – in this case the universal quantifier and implication –. The example of the cylinder showed that deforming "*A*" –of course *not* a formative sign– so that it came to include also the cylinder, entailed also a deformation of logical terms –in this case, the universal quantifier–. And Lakatos emphasizes that this was an important event, since it draws attention on the possibility that also logical notions experience some shifts of meaning.

Lakatos's brilliant insight, in this case, seems to foreshadow themes and procedures from J. Etchemendy's book *The Concept of Logical Consequence*,[18] where deformation of the logical constants is practised to the aim of reconciling Tarski's analysis of 1936 with informal notions, trying to escape from the danger of both undergeneration and overgeneration of logical truths. Just to give a simple example of this procedure, we remind that a consequence of the complex argument erected by Etchemendy against Tarski's analysis is that non-logical facts have no say in the process of assessing the truth-value of a sentence, *unless* non-logical expressions occur among the constant expression of the sentence. This assumption, however, is trivially false: facts concerning the cardinality of the domain of the structure are relevant to the truth value of a sentence even though it doesn't contain any expression of a non-logical type. Every sentence of the form "There exists at least *n* individuals", in fact, can be expressed by means of only existential quantifier, negation, and identity. For instance, $\exists x \exists y (x \neq y)$ doesn't contain any non-logical constant, and says that there exist at least two individuals. Since it can't be modified by reinterpreting the (not occurring) non-logical constants, if the sentence is true then it is dubbed logically true by the *Reduction Principle*.[19] And this, obviously, is absurd. To overcome this difficulty, and in this way banishing the possibility of overgeneration, Etchemendy *stretches*, we could say, the existential quantifier, formally including this logical notion into the set of the variable expressions. More precisely, he substitutes the existential quantifier "$\exists$" with a *quantificational*

[18] Etchemendy (1990).

[19] It is the principle according to which the *logical* truth of a given sentence *S* depends on the *ordinary* truth of the universal closure of the sentential function associated to *S*. For details see Mariani and Moriconi (1997).

variable "E" whose satisfaction domain consists of all subcollections of the universe of the structure (which, it is one of the fundamental assumptions made by Etchemendy, is fixed once and for all), instead of all individuals in the universe. Consequently, the previous sentence $\exists x \exists y(x \neq y)$ would be logically true iff $\forall E(Ex Ey(x \neq y)$ were ordinarily true. That is to say, iff every subcollection of the universe contained at least two elements. This last statement, however, is of course false: there are singletons. Thus, rightly, we get that $\exists x \exists y(x \neq y)$ isn't logically true.

Although empathic to an open idea of *deformation* of the concepts involved in the conjecture under attack, manoeuvres of the kind exploited by Etchemendy, open also to "a deformation of 'all' into 'no"' (p. 103), are not, however, what Lakatos was searching for. His aim, I think very well attained, was instead to bring out the link deeply connecting definitional procedures with the improving steps generated by the initial proof and the associated proof analysis (see points 1. and 2. of Sect. 13.2.1). All must be firmly grounded in the proof of the conjecture which is the real starting point of the growth and evolution of informal mathematical knowledge. By itself, the conjecture is not enough: it has to be supplied by a "proof", in the sense of point 1. of Sect. 13.2.1, which actually opens the way to the improvements of the initial "proof analysis" which lead possibly to a new proof of the original conjecture, but also to the possibility that we try giving up the conjecture and replace it by a new (possibly completely) different one.

13.5 Conclusion

Having outlined some of the remarkable issues from $\mathcal{P}\&\mathcal{R}$, I'd like lastly to consider a much less satisfactory position held by Lakatos, that is his firm belief that in *formal* mathematics there is no room for speaking of a growth of mathematical knowledge. In the already quoted (Lakatos, 1978), Lakatos provides a three-layered classification of mathematical proofs:

1. pre-formal proofs
2. formal proofs
3. post-formal proofs

and he stresses that proofs of type 1 and 3 are kinds of *informal* proofs. Relevant for our argument is the comment he adds in order to clarify the nature of post-formal proofs. He says that they fall under two types. The first one is exemplified by the Duality Principle in projective geometry,[20] the second one by the proofs of undecidability (and I think he meant to say "of *incompleteness* and undecidability"). He emphasizes that the Duality Principle works actually as a sort of meta-theorem,

[20] On a different plane, but equally pertinent, he could have mentioned other meta-theorems like the Deduction Theorem, the Löwenheim-Skolem Theorem or Gˊodel's Completeness Theorem.

or lemma, which allows to transform a theorem concerning "points" and "lines" in another one in which the words "point" and "line" are interchanged. The proof that brings us from the first to the second theorem, thanks to the Duality Principle, is an argument which cannot reach its completion without specifying the concepts of "provability" in the relevant system, of "theorem" in the system and so on. In other words, by exploiting something like the Duality Principle we don't prove a proposition which concerns *just* lines or points, but also points, lines, *provability*, *theoremhood* and so on.[21]

Turning to the second type of proofs, the proofs of undecidability, I rate opportune to stress that Lakatos overlooks the fact that also Gödel's Incompleteness Theorems "play a double role". On the one hand, indeed, they are *meta*-theorems in which, besides number-theoretical expressions, also meta-theoretical notions as substitution, theoremhood, ... are involved. On the other hand, however, thanks to the arithmetization of the meta-theory[22] they are also plain *theorems*. Lakatos conflates this kind of proofs and the so-called Gödel's proof that his undecidable sentences is *true* (i.e. true in the standard model). But, he adds,

> such post-formal proofs are certainly informal and so they are subject to falsification by the later discovery of some not-thought-of possibility.

Admittedly, this remark is rather puzzling: actually, the latter proof rightly comes under the heading "informal proof", but this same categorization is hardly fitting for the other mentioned (meta-)theorems. I think that what makes Lakatos's classification not completely satisfactory is the fact that it leads to conflate the meaning of "axiomatization" and "formalization" (as it is explicitly declared at p. 67 of Lakatos (1978)). Elsewhere, working on the basis of widespread ideas, I supported a representation of mathematical practice within a three-level framework:

1. informal, or pre-formal, mathematics,
2. (informal) axiomatic theories, and
3. formal theories.

where it is to be stressed that the three phases do not delete each other; all of them, so to speak, *live together*. Formal theories cannot be studied separately from all the non-formal, or pre-formal, background behind them.[23] The step from the second level to the third one reflects of course the remarkable change occurred between the end of nineteenth century and the Thirties of the twentieth century, when formalization came to the fore. Frege's and other people's discovery of the possibility to formalize mathematical knowledge produced in fact a new theoretical subject: *formal theories*. Here, "formal" extends its meaning, since

[21] And analogously, in the other results mentioned, not just formulas, but derivability or satisfiability of (set of) formulas.

[22] Also called *gödelization*.

[23] See Moriconi (2018).

- On the one hand, it means that a given part of mathematical knowledge undergoes a process of rebuilding which aims at making explicit its logical structure so that the justification procedure of informal mathematical proofs can be fully understood.
- On the other hand, "formal" means that we give up using natural languages and adopt conventional artificial ones, without any particular meaning assigned to (strings of) the signs of the language.

Axiomatization is distinct from formalization, where the latter notion refers to the second meaning of "formal"; and this, of course, even though axiomatization is a *necessary* condition of formalization, and even though any axiomatization exploits a (previous) work of formalization (referring in this case to the first sense of "formal").

However, it would be the outcome of a misunderstanding to consider the construction of a formal theory as a simple work of *decoration*. Indeed, a formal theory produces new questions, new knowledge, and new standpoints: in a word, and contrary to what Lakatos was inclined to believe, it produced and produces a rich growth of mathematical knowledge. The third layer marks the point where the system, beyond being the framework where investigation is developed, becomes an "object" of study in itself. First of all, in fact, without such notion of a formal system it would have been impossible to get the incompleteness results, and also to develop a formal theory of the way in which *truth* can be attributed to a sentence of a formal mathematical theory. Moreover, one could wonder whether when a *formal* proof of a mathematical sentence is available, we gain some knowledge which goes beyond knowing that sentence is provable (in a pre-formal theory, for instance). We can for instance argue about, and elaborate on, both the proof of a sentence –by providing a *normal* proof, or getting a speed-up, or extracting pieces of computational data from the proof of a sentence of first-order Peano Arithmetic– and its provability.

Lakatos does not like this kind of speculations, and I think that this fact doesn't depend only on his inclining to empiricism. It is deeply rooted in the actual difficulties he met in devising a program that I would like to call of "informal rigor"; that is, a program incorporating the idea of providing precise analysis of notions implicit in *common* mathematical reasoning and practice by trapping them between *formal* notions.[24] True, the aim of an *open framework*[25] doesn't fade away, even if we have to take into account the fact that, embedded into the text of the dialogue, there is the attempt to reach what we have called "The Logic of $\mathcal{P}\&\mathcal{R}$", and character Lambda (with subsequent integrations by Omega and Zeta) presents the "official

[24] I am of course borrowing the term, and the idea, from G. Kreisel's paper of 1967, *Informal Rigor and Completeness Proofs*, which was published in the volume *Problems in the Philosophy of Mathematics*, edited by I. Lakatos. In the first 50s, Kreisel had developed the no-counterexample interpretation of Peano Arithmetic: for interesting remarks on this point, involving also Lakatos's $\mathcal{P}\&\mathcal{R}$, see Fichot (2012a,b).

[25] See Sect. 13.2.1.

statement" of the heuristic rules of the method of proofs and refutations.[26] But the project remains in need of a completion.

First, because of a loose formulation which makes it difficult to evaluate the rules: Rule 2, for instance, demands

> to add to the proof-analysis a *suitable* lemma that will be refuted by the counterexample [my emphasis].

The occurrence of the adjective "suitable" stresses that the added lemma must be true to the spirit of the relevant conceptual experiment, so that ad hoc and casual conjectures are excluded: this caveat, however, is proposed without providing any hint of how to ascertain that the goal will be, or has been, attained.

Second, the project is weakened by a not completely satisfactory characterization of the notion of "formal", and by the firm belief that formal methods are anyway absolutely fruitless, so that no sort of dialectic between informal and formal notions is pursued. A consequence of these contrasting motivations is the substantial inability to produce new interesting lines of research: among the examples Lakatos makes it is worth remembering the possibility to *falsify* "the Zermelo-Fraenkel and kindred systems of formalized set theory". The argument seems to develop through the following steps.

1. Gödel had the firm *opinion* of the falsity of Cantor's continuum-hypothesis.
2. Gödel proved the *(meta-)theorem* asserting that (the Axiom of Choice and) Cantor's continuum-hypothesis is (are) consistent relatively to the Zermelo-Fraenkel and kindred systems of formalized set theory.
3. Thus, it is impossible to disprove in them Cantor's continuum-hypothesis.
4. Therefore, those systems are *falsified*.

Note: they are not rated to be unsuitable and in need of improvements. They are falsified.

In my opinion, the harshness of this conclusion, together with the already mentioned inability to produce new interesting lines of research while remaining within the same framework, are responsible for his shift from the history and philosophy of mathematics to the history and philosophy of the empirical sciences. In 1965 he organized in London an *International Colloquium in the Philosophy of Science*, which was an epoch-making event. The most famous of the "Conference Proceedings", *Criticism and the Growth of Knowledge*, contains his paper *Falsification and the Methodology of Scientific Research Programmes* which marks truly a turning point in the perspectives of his epistemological investigations.

[26] See Lakatos (1976), pp. 50, 58, and 76.

References

Bar-Am, N. 2009. Proof versus sound inference. In *Rethinking popper*, ed. Z. Parusniková and R.S. Cohen. Springer.

Binder, D., and T. Piecha. 2017. Popper's notion of duality and his theory of negation. *History and Philosophy of Logic* 38(2): 154–189.

Etchemendy, J. 1990. *The concept of logical consequence*. Harvard University Press.

Fichot, J. 2012a. Lakatos et l'interprétation sans contre-exemple. Extended Abstract of Communication at a Congress.

Fichot, J. 2012b. Preuves, réfutations et contre-exemples. In *IVe Congrès de la Société de Philosophie des Sciences*.

Kripke, S. 1982. *Wittgenstein on rules and private language*. Oxford: Blackwell.

Lakatos, I. 1976. *Proofs and refutations*. Cambridge University Press, Cambridge. The articles were originally published in the *British Journal for the Philosophy of Science*, 1963–1964.

Lakatos, I. 1978. What does a mathematical proof prove? In *Mathematics, Science and Epistemology: Philosophical Papers, vol. II*.

Mariani, M., and E. Moriconi. 1997. Etchemendy on logical truth. *Epistemologia* XX: 267–296.

Moriconi, E. 2018. Some remarks on *True* undecidable sentences. In *Truth, Existence and Explanation*, number 334 in Boston Studies in the Philosophy and History of Science, ed. M. Piazza and G. Pulcini, 3–15. Springer.

Moriconi, E. 2019. On Popper's decomposition of logical notions. In *Third Pisa colloquium in logic, language and epistemology*, ed. E.A. Luca Bellotti, 275–301. Edizioni Ets, Pisa.

Popper, K.R. 1946. Why are the calculuses of logic and arithmetic applicable to reality? *Arist Soc Supp* XX: 40–60.

Popper, K.R. 1947. Logic without assumptions. *Proceedings of the Artist in Society* 47: 251–292.

Tarski, A. 2002. On the concept of following logically. *History and Philosophy of Logic* 3: 155–196.

Chapter 14
A Categorical Reading of the Numerical Existence Property in Constructive Foundations

Samuele Maschio

Abstract We propose here an analysis based on syntactic categories and internal categories of existence properties. These metamathematical properties are peculiar of constructive theories, since they bring the internal notion of existence back to the external one, in accordance with the informal paradigm for constructivism known under the acronym BHK. Category theory is a powerful tool to analyse this phenomenon, since a category is an environment which allows to describe effectively internal and external notions and their relationship.

Keywords Constructivism · Internal categories · Existence properties

14.1 Existence in Constructive Mathematics

In his "*A Constructive Manifesto*", chapter 1 section 3 p.11 in Bishop and Bridges (1985), Bishop was clear about his view on existence in mathematics:

> Constructive existence is much more restrictive than the ideal existence of classical mathematics. The only way to show that an object exists is to give a finite routine for finding it, whereas in classical mathematics other methods can be used.

This view is clearly in contrast with a majority formalist view, defended e.g. by Poincaré ((1906), troisième article, III, and Troelstra and van Dalen (1988) p.19), according to which

> Existence can mean only one thing: freedom from contradiction.

or by Hilbert (in a letter to Frege, see e.g. p.69 of Shapiro (2005)):

> if the arbitrarily given axioms do not contradict one another with all their consequences, then they are true and the things defined by them exist.

S. Maschio (✉)
Dipartimento di Matematica "Tullio Levi-Civita", Padova, Italy
e-mail: maschio@math.unipd.it

G. Oliveri et al. (eds.), *Objects, Structures, and Logics*, Boston Studies in the Philosophy and History of Science 339,
https://doi.org/10.1007/978-3-030-84706-7_14

Indeed, this mainstream attitude directly leads to the identification of $\exists x\, P(x)$ with $\neg\forall x \neg P(x)$ and $\neg\neg\exists x\, P(x)$, which is exactly one of the methods to which Bishop referred to in the quotation above; the position of constructivism towards these methods is very well expressed by Bridges' words in Bridges (2008), section 3.1:

> how could a proof of the impossibility of the non-existence of a certain object x describe a mental construction of x?

This strong philosophical (and methodological) constructive view on existence can be translated in formal mathematical and metamathematical terms, as we will see in the next sections.

It is sort of intuitive to understand that the constructive notion of existence is captured by mathematical foundational theories for which the distance between the mathematical level and the metamathematical one (in which the first is defined) is minimized. Here we will use the descriptive and expressive power of category theory to illustrate this fact in a more structured way: we will move from syntax and mathematical theories to categories and internal categories, respectively, adopting a methodology similar (to some respects) to that of algebraic set theory (see e.g. Simpson 1999 or Maschio 2015).

14.2 A Paradigm for Constructive Proofs: BHK

In most textbooks about constructive mathematics (e.g. in Troelstra and van Dalen (1988), chapter 1 section 3 p.9 and in Bridges and Vîţă (2006), chapter 1 section 1.1. p.3), the underlying logical system is explained by means of an informal interpretation of what is a constructive proof of a compound formula, known under the name of BHK.[1] According to this interpretation:

- ($\wedge_{\mathsf{BHK}}$) a proof of $P \wedge Q$ is a pair $\langle p, q\rangle$ with p a proof of P and q a proof of Q;
- ($\vee_{\mathsf{BHK}}$) a proof of $P \vee Q$ is a pair $\langle i, r\rangle$ consisting of a proof r and a label i declaring whether that proof is a proof of P or a proof of Q;
- ($\rightarrow_{\mathsf{BHK}}$) a proof of $P \rightarrow Q$ is a procedure f turning proofs p of P into proofs $f(p)$ of Q;
- ($\exists_{\mathsf{BHK}}$) a proof of $\exists x\, P(x)$ is a pair $\langle a, p\rangle$ consisting of an object a and a proof p of $P(a)$;
- ($\forall_{\mathsf{BHK}}$) a proof of $\forall x\, P(x)$ is a procedure f which associates to each object a a proof $f(a)$ of $P(a)$.

This interpretation does not say what is a proof of an atomic formula and it is clearly informal. However, as we will see in the next two sections, there are at least two ways to make it "formal": one is syntactical, the other semantical.

[1] The acronym BHK comes from the names of three mathematicians which contributed to the constructive approach to mathematics, namely Brouwer, Heyting and Kolmogorov.

14.3 A Semantic Counterpart of BHK

In order to concretely accomplish BHK, proofs should enjoy at least these two properties: a pair of proofs must be a proof and some (partial) functions sending proofs into proofs must be proofs. Mathematically, we can render these requirements as follows: if $\mathbb{P}$ is a collection of proofs in BHK sense, then there must be an injective pairing function $\mathsf{pair} : \mathbb{P} \times \mathbb{P} \to \mathbb{P}$, and $\mathbb{P}$ must be endowed with a (partial) function from $\mathbb{P}$ to the set of partial functions from $\mathbb{P}$ to $\mathbb{P}$.

There is a meaningful structure validating these requirements: natural numbers. In fact there is a primitive recursive bijective encoding of pairs of natural numbers by means of natural numbers $(p : (n, m) \mapsto 2^n(2m + 1) - 1)$ with projections p_1 and p_2, and every natural number n represents a recursive (partial) function $\{n\}$ from $\mathbb{N}$ to $\mathbb{N}$, whenever a Gödelian encoding is fixed.

Using this structure on natural numbers, one can define the so-called Kleene realizability which is a rigorous semantical account of BHK for Heyting arithmetics, where proofs are interpreted as natural numbers. Kleene's realizability relation, which is represented by a formula $x \Vdash P$ ("x realizes P"), is defined by induction on the complexity of the formula P:[2]

$x \Vdash P \equiv^{def} P$ for atomic formulas P;

$(\wedge_{real})$ $x \Vdash P \wedge Q \equiv^{def} p_1(x) \Vdash P \wedge p_2(x) \Vdash Q$, that is, a realizer for $P \wedge Q$ is a natural number encoding a pair of natural numbers in which the first component realizes P and the second realizes Q;

$(\vee_{real})$ $x \Vdash P \vee Q \equiv^{def} (p_1(x) = 0 \wedge p_2(x) \Vdash P) \vee (p_1(x) = 1 \wedge p_2(x) \Vdash Q)$, that is, a realizer for $P \vee Q$ is a natural number enconding a pair in which the first component is a label which tells whether the second component is a realizer of P or a realizer of Q;

$(\to_{real})$ $x \Vdash P \to Q \equiv^{def} \forall y(y \Vdash P \to \{x\}(y) \downarrow \wedge \{x\}(y) \Vdash Q)$, that is, a realizer of $P \to Q$ is a code of a (partial) recursive function sending realizers of P to realizers of Q;

$(\exists_{real})$ $x \Vdash \exists z\, P \equiv^{def} p_2(x) \Vdash P[p_1(x)/z]$, that is, a realizer of $\exists z\, P$ is a natural number encoding a pair in which the second component is a realizer of the formula obtained from P by substituting z with the first component;

$(\forall_{real})$ $x \Vdash \forall z\, P \equiv^{def} \forall z(\{x\}(z) \downarrow \wedge \{x\}(z) \Vdash P)$, that is, a realizer of $\forall z\, P$ is a code of a total recursive function sending each natural number n to a realizer of $P[n/z]$.

Using Kleene realizability in Kleene (1945), in Troelstra (1971) it was proved that $\mathsf{HA} \vdash \exists x(x \Vdash \varphi) \Leftrightarrow \mathsf{HA} + \mathsf{ECT}_0 \vdash \varphi$, where HA is Heyting arithmetic and ECT_0 is the so-called *Extended Church's Thesis* (see Troelstra and van Dalen (1988),

[2] We always assume x and y not to occur in the formulas of which the realizability relation is defined.

chapter 4 section 4, p.199). In particular this provides a relative consistency proof of

$$\mathsf{HA} + \text{“All definable functions between natural numbers are computable”}$$

with respect to HA.[3]

There are many other notions of realizability arising from similar algebraic structures, which are called *partial combinatory algebras* (see e.g. Van Oosten (2008), chapter 1).

It should also be noticed that there exist in literature Kleene realizability models for intuitionistic set theories like Intuitionistic Zermelo-Fraenkel set theory IZF (see e.g. Friedman 1973, Rosolini 1982 and McCarty 1986); however in these cases one need to modify the interpretations of primitive formulas and quantifiers; in particular, since a realizer is a natural number, one cannot incorporate a witness for an existential statement (which would be a set) into it.

14.4 A Syntactic Counterpart of BHK

Per Martin-Löf introduced his intuitionistic type theory (see Martin-Löf 1984, Nordström et al. 1990) in the early 70's. From the first lines of p.1 in Martin-Löf (1975) one can understand his goal and the particular attention dedicated to the meaning of existential statements:

> The theory of types with which we shall be concerned is intended to be a full scale system for formalizing intuitionistic mathematics as developed, for example, in the book by Bishop. The language of the theory is richer that the languages of traditional intuitionistic systems in permitting proofs to appear as parts of propositions so that the propositions of the theory can express properties of proofs (and not only individuals, like in first order predicate logic). This makes it possible to strengthen the axioms for existence, disjunction, absurdity and identity. In the case of existence, this possibility seems first to have been indicated by Howard, whose proposed axioms are special cases of the existential elimination rule of the present theory.

Concretely, in Martin-Löf type theory a dependent sum type constructor Σ is defined by the following four rules:

1. a formation rule

$$\frac{A\ type \qquad B(x)\ type\ [x \in A]}{(\Sigma x \in A)B(x)\ type}$$

[3] The statement “All definable functions between natural numbers are computable” is not expressible as a formula in HA, but as a collection of formulas

$$\forall x \exists! y\, \varphi(x, y) \to \exists e \forall x (\varphi(x, \{e\}(x)))$$

for $\varphi(x, y)$ formula of HA.

which states that one can form the dependent sum $(\Sigma x \in A)B(x)$ of a family $B(x)$ of types indexed over a type A;

2. an introduction rule

$$\frac{a \in A \qquad b \in B(a)}{\langle a, b\rangle \in (\Sigma x \in A)B(x)}$$

which states that all pairs $\langle a, b\rangle$ with $a \in A$ and $b \in B(a)$ are terms of type $(\Sigma x \in A)B(x)$;

3. an elimination rule

$$\frac{\begin{array}{l} d \in (\Sigma x \in A)B(x) \\ C(z)\, type\, [z \in (\Sigma x \in A)B(x)] \\ c(x, y) \in C(\langle x, y\rangle)[x \in A, y \in B(x)] \end{array}}{\mathsf{El}_{\Sigma}(d, c(x, y)) \in C(d)}$$

which essentially says that nothing else is in $(\Sigma x \in A)B(x)$ (in order to assign an element of $C(d)$ to each $d \in (\Sigma x \in A)B(x)$, it is sufficient to assign an element of $C(\langle x, y\rangle)$ to each $x \in A$ and $y \in B(x)$);

4. an equality rule

$$\frac{\begin{array}{l} a \in A \qquad b \in B(a) \\ C(z)\, type\, [z \in (\Sigma x \in A)B(x)] \\ c(x, y) \in C(\langle x, y\rangle)[x \in A, y \in B(x)] \end{array}}{\mathsf{El}_{\Sigma}(\langle a, b\rangle, c(x, y)) = c(a, b) \in C(\langle a, b\rangle)}.$$

One of the key features of Martin-Löf type theory is the so called "propositions-as-types" paradigm: logic and mathematics are identified. For example, the dependent sum type Σ is used to represent the existential quantifier $\exists$, which, as a consequence, satisfies the following rules, which are obtained or derived from the rules above, by reading some types P as propositions and by interpreting the relative judgements of the form $p \in P$ as "p is a proof of P".

$$\frac{A\, type \qquad P(x)\, prop\, [x \in A]}{(\exists x \in A)P(x)\, prop} \qquad \frac{a \in A \qquad b \text{ is a proof of } P(a)}{\langle a, b\rangle \text{ is a proof of } (\exists x \in A)P(x)}$$

$$\frac{d \text{ is a proof of } (\exists x \in A)P(x)}{\pi_1(d) := \mathsf{El}_{\Sigma}(d, x(x, y)) \in A} \qquad \frac{d \text{ is a proof of } (\exists x \in A)P(x)}{\pi_2(d) := \mathsf{El}_{\Sigma}(d, y(x, y)) \text{ is a proof of } P(\pi_1(d))}$$

Hence, in Martin-Löf type theory the identification between Σ and $\exists$ imposes the validity of the request about existential statements in BHK. Other constructors of Martin-Löf type theory are designed in order to accomplish BHK via the propositions-as-types paradigm.

In other type theories, like in the Minimalist Foundation **MF** (see Maietti 2009; Maietti and Sambin 2005), which was introduced in order to provide a core foundation compatible with the most relevant classical and intuitionistic, predicative and impredicative, foundations, the paradigm propositions-as-types is not adopted. In the formulation of **MF** there is a distinction between two kinds of types: logical (propositions and small propositions) and mathematical (collections and sets) and the existential propositions satisfy rules which are similar to those of Σ-types above; however the elimination rule works only toward propositions. Hence one cannot in general produce the witness required by BHK. However one can show that **MF** admits a Kleene realizability interpretation (see Ishihara et al. 2018; Maietti and Maschio 2015).

14.5 Existence Properties

In first-order theories the requirement on existential quantifiers from BHK cannot be imposed in the formulation of the theory itself, as it is in fact done in Martin-Löf type theory. However, it can be controlled a posteriori, after being reformulated as a metamathematical property. The point is to find the right metamathematical formulation. In the literature there are many proposals; the difference between them consists in what they consider a *witness* for an existential statement should be.

In the first case witnesses are definable entities in the theory.

Definition 5.1 A first-order theory with equality $\mathcal{T}$ has the *existence property* (**EP**) if whenever $\mathcal{T} \vdash \exists x\ P(x)$, there exists a formula $Q(x)$, such that

$$\mathcal{T} \vdash \exists! x\ Q(x) \wedge \forall x (Q(x) \rightarrow P(x)).$$

The existence property **EP** essentially means that if something satisfying a property is proven to exist in $\mathcal{T}$, then something definable in $\mathcal{T}$ can be proven to satisfy that property.

The intuitionistic set theory IZF (see Friedman and Ščedrov 1985) and the constructive Zermelo-Fraenkel set theory CZF (see Swan 2014) do not have **EP**, while, as we will see in a few lines, Heyting arithmetic HA has it. Classical theories like Peano arithmetic PA and ZF+V=L, that is Zermelo-Fraenkel set theory with the additional axiom that states that all sets are constructible, also have **EP**. If $\mathsf{PA} \vdash \exists x\ P(x)$, then one can take $Q(x)$ to be $P(x) \wedge \forall y (P(y) \rightarrow x \leq y)$ which works because of the minimum principle which is provable in PA. In ZF+V=L one can do essentially the same, because there one can define a well-ordering on the universe class V.

Although at first sight **EP** could be considered a good candidate to express the BHK requirement about existential quantifiers, one could object that the "unique existence" required in the definition could be proven in $\mathcal{T}$ by means of indirect methods, thus producing a *witness* being only apparently "concrete".

Another option could be to consider a *term existence property* in which witnesses are simply terms not containing variables. However this is not of great interest in this framework:[4] terms representing definable objects can indeed be added to a first-order theory leaving it essentially equivalent and turning, in the end, term existence property into existence property.

Looking for something sufficiently simple to be considered "stable" from the external point of view, one comes to numerals, that is natural (meta)numbers. In fact the notion of numeral requires only the ability of juxtaposing symbols, which is a minimal requirement for being able to formulate a first-order theory. In this sense we can think of numerals as a good notion of witnesses. However they can only be used as witnesses for those formulas in which the free variable represents a natural number in the sense of the theory $\mathcal{T}$:

Definition 5.2 A first-order theory of natural numbers $\mathcal{T}$ has the *numerical existence property* (**nEP**) if, for every formula $P(x)$,[5] there exists a numeral $\mathbf{n}$ such that $\mathcal{T} \vdash P(\mathbf{n})$, whenever $\mathcal{T} \vdash \exists x\, P(x)$. $\mathcal{T}$ has the *unique numerical existence property* (**nEP**$_!$) if, for every formula $P(x)$, there exists a numeral $\mathbf{n}$ such that $\mathcal{T} \vdash P(\mathbf{n})$, whenever $\mathcal{T} \vdash \exists! x\, P(x)$.

A first-order theory of sets $\mathcal{T}$, in which the existence of the set ω is provable, has the *numerical existence property* (**nEP**) if, for every formula $P(x)$, there exists a numeral $\mathbf{n}$ such that $\mathcal{T} \vdash P(\mathbf{n})$, whenever $\mathcal{T} \vdash \exists x \in \omega\, P(x)$. $\mathcal{T}$ has the *unique numerical existence property* (**nEP**$_!$) if, for every formula $P(x)$, there exists a numeral $\mathbf{n}$ such that $\mathcal{T} \vdash P(\mathbf{n})$, whenever $\mathcal{T} \vdash \exists! x \in \omega\, P(x)$.[6]

The numerical existence property **nEP** essentially means that if a natural number satisfying a property is proven to exist in $\mathcal{T}$, then a numeral can be proven to satisfy that property. **nEP**$_!$ essentially means that definable natural numbers exactly coincide with numerals.

Peano arithmetic PA, Zermelo-Fraenkel set theory ZF and, in general, classical first-order theories $\mathcal{T}$ of numbers or sets (if consistent) do not have the numerical existence property, not even the unique one. Indeed one can consider an independent sentence I (which exists by Gödel's first incompleteness theorem): clearly $\mathcal{T} \vdash \exists x((x = 0 \wedge \neg I) \vee (x = 1 \wedge I))$ as a consequence of the law of excluded middle; however there cannot be a numeral $\mathbf{n}$ such that $\mathcal{T} \vdash (\mathbf{n} = 0 \wedge \neg I) \vee (\mathbf{n} = 1 \wedge I)$, since in that case $\mathbf{n}$ would be 0 or 1 and we could hence prove $\neg I$ or I in $\mathcal{T}$. Heyting arithmetic HA has the numerical existence property: this was proven by means of realizability by Kleene (see Kleene 1945). CZF and IZF also have the numerical existence property, as was proven in Rathjen (2005) Theorem 1.2. and in Beeson

[4] The internal language of a doctrine in category theory is an example of framework in which terms have a clear "stable" meaning, that is "arrows of the base category", and where, hence, term existence property would be meaningful.

[5] When we write a formula $P(x_1, \ldots, x_n)$ we mean that P contains at most $x_1, \ldots, x_n$ as free variables.

[6] In set theory, $P(\mathbf{n})$ is defined as follows: $P(0) \equiv^{def} \exists x(\forall y(y \notin x) \wedge P(x))$ and for every natural (meta)number n, $P(\mathbf{n+1}) \equiv^{def} \exists x(\forall y(y \in x \leftrightarrow y \in \mathbf{n} \vee y = \mathbf{n}) \wedge P(x))$.

(1985) chapter VIII section 9, respectively. Clearly, for first-order theories of natural numbers $\mathbf{nEP} \equiv \mathbf{EP} + \mathbf{nEP}_!$ (hence HA has **EP**), while for first-order theories of sets $\mathbf{EP} + \mathbf{nEP}_! \Rightarrow \mathbf{nEP}$, but the converse does not necessarily hold.

14.6 Categories of Definable Classes

The next step consists in organizing the content of a first-order theory with equality in a category of definable classes and to introduce some useful subcategories.

We define the category $\mathsf{DC}[\mathcal{T}]$ of the definable classes of $\mathcal{T}$ as follows:

1. we first fix two variables x, y;
2. the objects of $\mathsf{DC}[\mathcal{T}]$ are formal expressions $\{x|\ P(x)\}$ where $P(x)$ is a formula; we identify objects $\{x|\ P(x)\}$ and $\{x|\ Q(x)\}$ with $P(x)$ and $Q(x)$ provable to be equivalent in $\mathcal{T}$;
3. an arrow from $\{x|\ P(x)\}$ to $\{x|\ Q(x)\}$ is a formula $F(x, y)$, such that
   (a) $F(x, y) \vdash_{\mathcal{T}} P(x) \wedge Q(y)$;
   (b) $F(x, y) \wedge F(x, z) \vdash_{\mathcal{T}} y = z$ (where z is a fresh variable);
   (c) $P(x) \vdash_{\mathcal{T}} \exists y F(x, y)$;

   and we identify formulas provable to be equivalent in $\mathcal{T}$;
4. the composition $G(x, y) \circ F(x, y)$ is defined as $\exists z(F(x, z) \wedge G(z, y))$ where z is a fresh variable;
5. the identity arrow of an object $\{x|\ P(x)\}$ is defined as the formula $P(x) \wedge x = y$.

For the theory $\mathcal{T}$ we can also define a category $\mathsf{DC}_{term}[\mathcal{T}]$ having the same objects as $\mathsf{DC}[\mathcal{T}]$, but for which an arrow from $\{x|\ P(x)\}$ to $\{x|\ Q(x)\}$ is an equivalence class $[t(x)]_{\simeq_{P(x)}}$ of terms $t(x)$,[7] such that $P(x) \vdash_{\mathcal{T}} Q(t(x))$ with respect to the relation $\simeq_{P(x)}$ for which $t(x) \simeq_{P(x)} s(x)$ when $P(x) \vdash_{\mathcal{T}} t(x) = s(x)$; the composition $[s(x)]_{\simeq_{Q(x)}} \circ [t(x)]_{\simeq_{P(x)}}$ of two arrows is defined by $[s(t(x))]_{\simeq_{P(x)}}$, while the identity $\mathsf{id}_{\{x|\ P(x)\}}$ is given by $[x]_{\simeq_{P(x)}}$.

The category $\mathsf{DC}_{term}[\mathcal{T}]$ is clearly a subcategory of $\mathsf{DC}[\mathcal{T}]$: just consider the functor sending each $\{x|\ P(x)\}$ to itself and each $[t(x)]_{\simeq_{P(x)}}$ to $P(x) \wedge y = t(x)$.

If $\mathcal{T}$ is a theory of natural numbers having at least 0 and the successor symbol s as primitive function symbols, we denote with $\mathsf{DC}_{nat}[\mathcal{T}]$ the subcategory of $\mathsf{DC}_{term}[\mathcal{T}]$ which have the same objects, the same definitions of composition and identity, but only those arrows which are representable by terms obtained using the variable x and the function symbols 0 and s.

If $\mathcal{T}$ is a theory of sets, we first translate terms τ of the language obtained using the variable x and function symbols 0 and s into formulas $[[\tau]]$ of $\mathcal{T}$ as follows: $[[0]] \equiv^{def} \forall z\,(z \notin y)$, $[[x]] \equiv^{def} x = y$ and $[[\mathsf{s}(\tau)]] \equiv^{def} \exists z\,([[\tau]][z/y] \wedge \forall u(u \in y \leftrightarrow u = z \vee u \in z))$.

[7] We will write $t(x_1, \ldots x_n)$ if the term t contains at most $x_1, \ldots, x_n$ as variables.

The objects of $\mathsf{DC}_{nat}[\mathcal{T}]$ are defined as those objects $\{x|\, P(x)\}$ of $\mathsf{DC}[\mathcal{T}]$ such that $P(x) \vdash_{\mathcal{T}} x \in \omega$; an arrow in $\mathsf{DC}_{nat}[\mathcal{T}]$ from $\{x|\, P(x)\}$ to $\{x|\, Q(x)\}$ is an arrow of $\mathsf{DC}[\mathcal{T}]$ of the form $[[\tau]] \wedge P(x)$ for some term τ of the language obtained using the variable x and function symbols 0 and s. One can prove that compositions and identities inherited from $\mathsf{DC}[\mathcal{T}]$ work with this restriction.

For every pair of objects $A = \{x|\, P(x)\}$ and $B = \{x|\, Q(x)\}$ in $\mathsf{DC}_{nat}[\mathcal{T}]$ an injection $J_{nat}^{A,B}$ can be defined as follows. If $\mathcal{T}$ is a theory of natural numbers:

$$J_{nat}^{A,B} : \mathsf{DC}_{nat}[\mathcal{T}](A, B) \to \mathsf{DC}[\mathcal{T}](A, B)$$

$$[t(x)]_{\simeq_P} \mapsto P(x) \wedge y = t(x)$$

If $\mathcal{T}$ is a theory of sets and A and B are objects of $\mathsf{DC}_{nat}[\mathcal{T}](A, B)$, $J_{nat}^{A,B}$ is the obvious inclusion of $\mathsf{DC}_{nat}[\mathcal{T}](A, B)$ in $\mathsf{DC}[\mathcal{T}](A, B)$.

If the theory $\mathcal{T}$ has at least a definable element and a definable encoding of ordered pairs, that is, if we assume that there exist a formula $I(x)$ such that $\mathcal{T} \vdash \exists! x I(x)$ and a formula $Pr(x, y, z)$ with three free variables x, y, z such that

1. $Pr(x, y, z) \wedge Pr(x, y, z') \vdash_{\mathcal{T}} z = z'$;
2. $Pr(x, y, z) \wedge Pr(x', y', z) \vdash_{\mathcal{T}} x = x' \wedge y = y'$;
3. $\vdash_{\mathcal{T}} \forall x \forall y \exists z\, Pr(x, y, z)$,

then $\mathsf{DC}[\mathcal{T}]$ is a cartesian category: a terminal object 1 is given by $\{x|\, I(x)\}$ and a product of $\{x|\, P(x)\}$ and $\{x|\, Q(x)\}$ is given by the object $\{x|\, \exists y \exists z (P(y) \wedge Q(z) \wedge Pr(y, z, x))\}$ together with the obvious projections.

This is the case for all standard theories of natural numbers and of sets: in HA or PA the formula $I(x)$ can be taken to be $x = 0$ and $Pr(x, y, z)$ to be $z = 2^x(2y+1)$;[8] in set theories like CZF, IZF and ZFC, $I(x)$ can be taken to be $\forall y (y \notin x)$ and $Pr(x, y, z)$ to be $z = \{\{x\}, \{x, y\}\}$.[9]

In the rest of the chapter we will always implicitely assume $\mathcal{T}$ to be a first-order classical or intuitionistic theory of sets or of numbers having at least a definable element and an encoding of ordered pairs.

14.7 Existence Properties, Categorically

We now show what properties of the categories introduced in the previous section correspond to the existential properties introduced in Sect. 14.5.

Before proving our characterization, let us recall some categorical notions: an arrow e in a category $\mathbb{C}$ is a *regular epi* if there exist arrows f and g in $\mathbb{C}$ of which e is the coequalizer, that is $e \circ f = e \circ g$ and for every arrow e' such that $e' \circ f = e' \circ g$,

[8] The exponential $2^{(-)}$ can be adequately represented by a definable relation.

[9] Here $z = \{x, y\}$ is an shorthand for $\forall u (u \in z \leftrightarrow (u = x \vee u = y))$.

there exists a unique arrow r such that $r \circ e = e'$; an arrow $e : A \to B$ in $\mathbb{C}$ is a *split epi* if there exists an arrow e' such that $e \circ e' = \mathsf{id}_B$.

Theorem 7.1 *Let $\mathcal{T}$ be a theory of natural numbers or a theory of sets as in the previous sections and let $P(x)$ be a formula of $\mathcal{T}$. Then*

1. *$\mathcal{T} \vdash \exists x\, P(x)$ if and only if the unique arrow from $\{x|\, P(x)\}$ to 1 in $\mathsf{DC}[\mathcal{T}]$ is a regular epi;*
2. *$\mathcal{T} \vdash \exists! x\, P(x)$ if and only if $I(x) \wedge P(y) : 1 \to \{x|\, Q(x)\}$ is a well-defined arrow in $\mathsf{DC}[\mathcal{T}]$ for every $Q(x)$ such that $P(x) \vdash_{\mathcal{T}} Q(x)$;*
3. *$\mathcal{T}$ has* **EP** *if and only if every regular epi in $\mathsf{DC}[\mathcal{T}]$ with codomain 1 is a split epi in $\mathsf{DC}[\mathcal{T}]$;*
4. *if $\mathcal{T}$ is a theory of natural numbers, $[t(x)]_{\simeq_{I(x)}} : 1 \to \{x|\, P(x)\}$ is an arrow in $\mathsf{DC}_{nat}[\mathcal{T}]$ if and only if there exists a numeral $\mathbf{n}$ such that $\mathbf{n} \simeq_{I(x)} t(x)$;*
5. *if $\mathcal{T}$ is a theory of sets and $[[\tau]] \wedge \forall u(u \notin x) : 1 \to \{x|\, P(x)\}$ is an arrow in $\mathsf{DC}_{nat}[\mathcal{T}]$, there exists a numeral $\mathbf{n}$ such that $[[\tau]] \wedge \forall u(u \notin x)$ and $[[\mathbf{n}]] \wedge \forall u(u \notin x)$ represent the same arrow from 1 to $\{x|\, P(x)\}$;*
6. *$\mathcal{T}$ has* **nEP** *if and only if every regular epi in $\mathsf{DC}[\mathcal{T}]$ from an object A to 1 is a split epi with right inverse of the form $J^{1,A}_{nat}(f)$ with f in $\mathsf{DC}_{nat}[\mathcal{T}]$;*
7. *$\mathcal{T}$ has* **nEP**$_!$ *if and only if $J^{1,A}_{nat}$ is a bijiection for every A in $\mathsf{DC}_{nat}[\mathcal{T}]$.*

Proof

1. Suppose that $\mathcal{T} \vdash \exists x\, P(x)$. Then one can prove that the unique arrow from $\{x|P(x)\}$ to 1 is the coequalizer of the two projections from $\{x|P(x)\} \times \{x|P(x)\}$ to $\{x|P(x)\}$. Conversely, suppose that the unique arrow ! from $\{x|P(x)\}$ to 1 is the coequalizer of two arrows f and g; then if we consider the unique arrow i from $\{x|P(x)\}$ to $\{x|I(x) \wedge \exists y\, P(y)\}$, then clearly $i \circ f = i \circ g$. In particular, it follows (since ! is the coequalizer of f and g) that there exists an arrow from 1 to $\{x|I(x) \wedge \exists y\, P(y)\}$; this entails that $\mathcal{T} \vdash \exists x\, P(x)$.
2. If $\mathcal{T} \vdash \exists! x\, P(x)$ and $P(x) \vdash_{\mathcal{T}} Q(x)$, then $I(x) \wedge P(y) \vdash_{\mathcal{T}} I(x) \wedge Q(y)$, $(I(x) \wedge P(y)) \wedge (I(x) \wedge P(z)) \vdash_{\mathcal{T}} y = z$ and $I(x) \vdash_{\mathcal{T}} \exists y(I(x) \wedge P(y))$; conversely, if $I(x) \wedge P(y) : 1 \to \{x|\, Q\}$ is a well-defined arrow in $\mathsf{DC}[\mathcal{T}]$ for every $Q(x)$ such that $P(x) \vdash_{\mathcal{T}} Q(x)$, then in particular $I(x) \vdash_{\mathcal{T}} \exists y(I(x) \wedge P(y))$ and $I(x) \wedge P(y) \wedge P(z) \vdash_{\mathcal{T}} y = z$; since $\mathcal{T} \vdash \exists x\, I(x)$, then $\mathcal{T} \vdash \exists! y\, P(y)$.
3. Suppose that $\mathcal{T}$ has **EP** and suppose that $P(x) \wedge I(y)$, which is the unique arrow from $\{x|\, P(x)\}$ to 1 in $\mathsf{DC}[\mathcal{T}]$, is a regular epi in $\mathsf{DC}[\mathcal{T}]$; then by point 1. we have that $\mathcal{T} \vdash \exists x\, P(x)$; as a consequence of **EP**, we have that there exists $Q(x)$ such that $\mathcal{T} \vdash \exists! x\, Q(x)$ and $Q(x) \vdash_{\mathcal{T}} P(x)$. These conditions together with point 2. allow to conclude that $I(x) \wedge Q(y)$ is a well-defined arrow from 1 to $\{x|P(x)\}$. Clearly this arrow is a right inverse of the unique arrow from $\{x|\, P(x)\}$ to 1. Conversely, suppose that $P(x)$ is a formula for which $\mathcal{T} \vdash \exists x\, P(x)$. By point 1. the unique arrow from $\{x|\, P(x)\}$ to 1 in $\mathsf{DC}[\mathcal{T}]$ is a regular epi, hence it is a split epi. This means that there is an arrow $F(x, y)$ from 1 to $\{x|P(x)\}$. One can see immediately that, since $\mathcal{T} \vdash \exists! x\, I(x)$, the formula $F(x, y)$ is equivalent in $\mathcal{T}$ to $I(x) \wedge \exists z\, F(z, y)$. If we take $Q(x)$ to be $\exists z(F(z, x))$, then the requirement of **EP**, applied to $P(x)$, is satisfied by $Q(x)$.

4. and 5. follow from the very definition of $\mathsf{DC}_{nat}[\mathcal{T}]$ and of numerals. From these, points 6. and 7. follow immediately. □

14.8 Internalizing $\mathsf{DC}[\mathcal{T}]$ in Itself

From now on, we will consider only theories $\mathcal{T}$ which enjoy a primitive recursive internal Gödelian encoding of their syntax by means of natural numbers. We also use, with abuse of notations, symbols for recursive function between natural numbers (including a primitive recursive bijective encoding of natural numbers p with primitive recursive projections p_1 and p_2), since they can be adequately represented in $\mathcal{T}$. In particular, every variable ξ in the syntax of $\mathcal{T}$ is encoded by a numeral $\boldsymbol{\xi}$. One can hence define a formula $\mathsf{dc}(x) \equiv^{def} \mathsf{form}(x) \wedge \forall y(\mathsf{free}(y, x) \rightarrow y = \mathbf{x})$ which expresses the fact that x is the code of a formula of $\mathcal{T}$ having at most x as free variable, in such a way that the definable class $\Delta\Gamma_0 := \{x | \mathsf{dc}(x)\}$, which is an object of $\mathsf{DC}[\mathcal{T}]$, is an internalization of the collection of objects of $\mathsf{DC}[\mathcal{T}]$ itself. However, in the definition of $\mathsf{DC}[\mathcal{T}]$, we have identified definable classes which were given by provably equivalent formulas. We hence need to take this into account internally by means of the obvious internal equivalence relation $\equiv_0$:

$$\{x | \exists y \exists z(x = \mathsf{p}(y, z) \wedge \mathsf{dc}(y) \wedge \mathsf{dc}(z) \wedge \mathsf{der}(y, z) \wedge \mathsf{der}(z, y))\} \rightarrow \Gamma\Delta_0 \times \Gamma\Delta_0$$

where $\mathsf{der}(x, y)$ is a formula expressing the fact that the formula encoded by y can be derived from that encoded by x in $\mathcal{T}$.

Analogously, one can define a formula $\mathsf{fr}(x)$ expressing the fact that x is the code of a definable functional relation of the form $F(x, y)$.

However, in order to encode the collection of arrows of $\mathsf{DC}[\mathcal{T}]$, we need to keep track of their codomains (which can not be reconstructed otherwise). We hence consider the collection

$$\Delta\Gamma_1 := \{x | \exists y \exists z(x = \mathsf{p}(y, z) \wedge \mathsf{fr}(y) \wedge \mathsf{dc}(z) \wedge \mathsf{der}(y, \mathsf{sub}(z, \mathbf{y}, \mathbf{x})))\}$$

(where $\mathsf{sub}(z, \mathbf{y}, \mathbf{x})$ is a term representing a code for the formula encoded by z in which the variable x is substituted with y after having renamed all occurrences of the variable y with a fresh variable primitively recursively depending on z) which is indeed an internal account of the collection of arrows of $\mathsf{DC}[\mathcal{T}]$ once we consider the obvious internal equivalence relation $\equiv_1$ with domain

$$\{x | \exists y \exists z(x = \mathsf{p}(y, z) \wedge y \varepsilon \Delta\Gamma_1 \wedge z \varepsilon \Delta\Gamma_1 \wedge$$

$$\mathsf{der}(\mathsf{p}_1(y), \mathsf{p}_1(z)) \wedge \mathsf{der}(\mathsf{p}_1(z), \mathsf{p}_1(y)) \wedge \mathsf{der}(\mathsf{p}_2(y), \mathsf{p}_2(z)) \wedge \mathsf{der}(\mathsf{p}_2(z), \mathsf{p}_2(y)))\}.$$[10]

[10] If $C = \{x | P(x)\}$ is a definable class, then we write $t \varepsilon C$ as a shorthand for $P(t)$.

In $\mathsf{DC}[\mathcal{T}]$ one can also define internally arrows corresponding to the domain, codomain, identity and composition operations (where we use the notation $\overline{-}$ to denote the Gödelian encoding in terms of primitive recursive functions of connectives, quantifiers and equality symbols):

1. the domain arrow is $\delta_0 := x \varepsilon \Delta\Gamma_1 \wedge y = \overline{\exists \mathbf{y}} \mathsf{p}_1(x) : \Delta\Gamma_1 \to \Delta\Gamma_0$;
2. the codomain arrow is $\delta_1 := x \varepsilon \Delta\Gamma_1 \wedge y = \mathsf{p}_2(x) : \Delta\Gamma_1 \to \Delta\Gamma_0$;
3. the identity arrow is $\mathsf{ID} := x \varepsilon \Delta\Gamma_0 \wedge y = \mathsf{p}(x \overline{\wedge} (\mathbf{x} \equiv \mathbf{y}), x) : \Delta\Gamma_0 \to \Delta\Gamma_1$;
4. the composition arrow $\Box : \mathsf{Pb}(\delta_1, \delta_0) \to \Delta\Gamma_1$, where $\mathsf{Pb}(\delta_1, \delta_0)$ denotes the obvious choice of a pullback for δ_1 and δ_0, is more complicated but can be easily formulated with some patience.

What really matters is that $\Delta\Gamma[\mathcal{T}] := ((\Delta\Gamma_0, \equiv_0), (\Delta\Gamma_1, \equiv_1), \delta_0, \delta_1, \mathsf{ID}, \Box)$ is essentially an internal category (see e.g. MacLane and Moerdijk (1992) chapter V section 7) in the elementary quotient completion of $\mathsf{DC}[\mathcal{T}]$ with respect to the doctrine of its subobjects (subobjects in $\mathsf{DC}[\mathcal{T}]$ can be represented by comprehension), since the arrows δ_0, δ_1, ID and $\Box$ respect the internal equivalence relations $\equiv_0$ and $\equiv_1$.

One can notice that in the case in which a particular *parametric version* of **EP** holds for formulas restricted to natural numbers (e.g. in classical set theory and in Peano arithmetic), one can avoid internal equivalence relations and choose representatives via a formula, obtaining an internal account of $\mathsf{DC}[\mathcal{T}]$ as one of its internal categories. More precisely, if $\mathsf{Nat}(x)$ is a formula in $\mathcal{T}$ expressing that x is a natural number, the particular parametric version of **EP** we consider is the following: whenever $P(x, y)$ is a formula with at most x and y as free variables such that $P(x, y) \vdash_{\mathcal{T}} \mathsf{Nat}(x) \wedge \mathsf{Nat}(y)$, there exists another formula $Q(x, y)$ with at most x and y as free variables such that

1. $Q(x, y) \vdash_{\mathcal{T}} P(x, y)$;
2. $\mathcal{T} \vdash \forall x (\exists y\, P(x, y) \to \exists! y\, Q(x, y))$;
3. $\forall y (P(x, y) \leftrightarrow P(x', y)) \wedge Q(x, z) \wedge Q(x', z') \vdash_{\mathcal{T}} z = z'$.

14.9 Numerical Existence Property and the Relation Between $\mathsf{DC}[\mathcal{T}]$ and $\Delta\Gamma[\mathcal{T}]$

In this section we recall a result in Maschio (2020) which connects internal categories and the numerical existence property.

First, we can observe that, whenever one has a internal equivalence relation $r : R \to I \times I$ in a finitely complete category $\mathbb{C}$, then the subset

$$\{(\pi_1 \circ r \circ f, \pi_2 \circ r \circ f) \mid f \in \mathsf{Hom}_{\mathbb{C}}(1, R)\} \subseteq \mathsf{Hom}_{\mathbb{C}}(1, I) \times \mathsf{Hom}_{\mathbb{C}}(1, I)$$

is an equivalence relation which we denote with $\mathsf{Ext}(r)$. Using this fact, one can define a category $\mathsf{Ext}(\Delta\Gamma[\mathcal{T}])$ "externalising" the internal category $\Delta\Gamma[\mathcal{T}]$ in $\mathsf{DC}[\mathcal{T}]$ as follows:

1. the collection of objects of $\mathsf{Ext}(\Delta\Gamma[\mathcal{T}])$ is $\mathsf{Hom}_{\mathsf{DC}[\mathcal{T}]}(1, \Delta\Gamma_0)/\mathsf{Ext}(\equiv_0)$;
2. the collection of arrows of $\mathsf{Ext}(\Delta\Gamma[\mathcal{T}])$ is $\mathsf{Hom}_{\mathsf{DC}[\mathcal{T}]}(1, \Delta\Gamma_1)/\mathsf{Ext}(\equiv_1)$;
3. the domain function Δ_0 is defined by $\Delta_0([f]) := [\delta_0 \circ f]$;
4. the codomain function Δ_1 is defined by $\Delta_1([f]) := [\delta_1 \circ f]$;
5. the identity function id is defined by $\mathsf{id}([f]) := [\mathsf{ID} \circ f]$;
6. the composition function $\circ$ is defined by $[g] \circ [f] := [\square \circ \langle f, g \rangle]$.[11]

In general one can define a functor $\mathbf{J} : \mathsf{DC}[\mathcal{T}] \to \mathsf{Ext}(\Delta\Gamma[\mathcal{T}])$ as follows:

1. If $\{x|\, P(x)\}$ is an object of $\mathsf{DC}[\mathcal{T}]$, then the formula $P(x)$ is encoded by a numeral $\mathbf{cod}(P(x))$ for which clearly $\mathcal{T} \vdash \mathsf{dc}(\mathbf{cod}(P(x)))$. We hence define $\mathbf{J}(\{x|\, P(x)\})$ as the equivalence class represented by the arrow

$$I(x) \wedge y = \mathbf{cod}(P(x)) : 1 \to \Delta\Gamma_0;$$

2. if $F(x, y) : \{x|\, P(x)\} \to \{x|\, Q(x)\}$ is an arrow in $\mathsf{DC}[\mathcal{T}]$, then $F(x, y)$ is encoded by a numeral $\mathbf{cod}(F(x, y))$ and $Q(x)$ is encoded by a numeral $\mathbf{cod}(Q(x))$ and, hence, $\mathcal{T} \vdash \mathsf{p}(\mathbf{cod}(F(x, y)), \mathbf{cod}(Q(x)))\varepsilon\Delta\Gamma_1$. We hence send $F(x, y) : \{x|\, P(x)\} \to \{x|\, Q(x)\}$ to the equivalence class represented by the arrow

$$I(x) \wedge y = \mathsf{p}(\mathbf{cod}(F(x, y)), \mathbf{cod}(Q(x))) : 1 \to \Delta\Gamma_1.$$

We can hence enunciate the following result which is proven in Maschio (2020) in Theorem 7.4

Theorem 9.1 $\mathbf{J}$ *is an isomorphism if* $\mathcal{T}$ *has the* **nEP**.

14.10 A Categorical Reading of Numerical Existence Property in Constructive Foundations

A category $\mathbb{C}$ with enough structure (e.g. a Heyting category or a topos) can be thought of as a mathematical universe in which one can perform "internal mathematics". The internal mathematics performed in different categories satisfy different principles and each category has its own internal groups, rings, preorders... As we have seen, categories can also have their own internal categories. These internal categories have a completely different nature than the category in which they live (e.g. trivially objects of the external category form a set or a class, while in general

[11] We denote with $\langle f, g \rangle$ the unique arrow determined by the definition of pullback.

the objects of objects of its internal categories need not be sets or classes). These two different levels of categories can be thought of as representations of the two levels corresponding to metamathematics and mathematics.

In our case the external category, $\mathsf{DC}[\mathcal{T}]$, is in fact defined in the metamathematical level: its objects are equivalence (meta)classes of formal expressions (forming a countable (meta)set) and the same holds for arrows. The internalization $\Delta\Gamma[\mathcal{T}]$ is an internal category of $\mathsf{DC}[\mathcal{T}]$ of which the objects do not form a (meta)set, although they form a class of elements from the point of view of the theory $\mathcal{T}$. We could roughly say that $\mathsf{DC}[\mathcal{T}]$ is an ordinary category from the point of view of the metamathematician, while $\Delta\Gamma[\mathcal{T}]$ is an ordinary category from the point of view of the mathematician working in the theory $\mathcal{T}$. We can think of $\Delta\Gamma[\mathcal{T}]$ as the best representation of the category $\mathsf{DC}[\mathcal{T}]$ that a mathematician working in $\mathcal{T}$ (and using only the tools of $\mathcal{T}$) can obtain. The metamathematician knows both the category $\mathsf{DC}[\mathcal{T}]$ and the internal category $\Delta\Gamma[\mathcal{T}]$ and he could ask himself whether from its point of view $\Delta\Gamma[\mathcal{T}]$ is a good representation of $\mathsf{DC}[\mathcal{T}]$. This, as we have seen, is done by means of a simulation of the notion of elements for objects of the category, that is by means of global elements.

Theorem 9.1 essentially means that whenever the numerical existence property is satisfied (which happens essentially only for some constructive theories), then the metamathematician can consider $\Delta\Gamma[\mathcal{T}]$ a perfect representation of $\mathsf{DC}[\mathcal{T}]$. This in some sense breaks some portions of the floor separating the mathematical level and the metamathematical one.

As we have already said, categorical language is not necessary to understand this, however it provides a clearer picture, a concrete representation in which syntactical aspects are organized in such a way that they can form structures which are more familiar to mathematicians.

References

Beeson, M.J. 1985. *Foundations of constructive mathematics*, volume 6 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Berlin: Springer. Metamathematical studies.

Bishop, E., and D.S. Bridges. 1985. *Constructive analysis*. Springer.

Bridges, D.S. 2008. Constructive mathematics. In *Stanford Encyclopedia of Philosophy*.

Bridges, D.S., and L.S. Vîţă. 2006. *Techniques of constructive analysis*. Universitext. New York: Springer.

Friedman, H.M. 1973. Some applications of Kleene's methods for intuitionistic systems. In *Cambridge summer school in mathematical logic (Cambridge, 1971)*, Lecture notes in mathematics, Vol. 337, 113–170.

Friedman, H.M. and A. Ščedrov. 1985. The lack of definable witnesses and provably recursive functions in intuitionistic set theories. *Advances in Mathematics* 57(1): 1–13.

Ishihara, H., M.E. Maietti, S. Maschio, and T. Streicher. 2018. Consistency of the intensional level of the minimalist foundation with Church's thesis and axiom of choice. *Archive for Mathematical Logic* 57(7–8): 873–888.

Kleene, S.C. 1945. On the interpretation of intuitionistic number theory. *The Journal of Symbolic Logic* 10(4): 109–124.

MacLane, S., and I. Moerdijk. 1992. *Sheaves in geometry and logic. A first introduction to Topos theory.* Springer.

Maietti, M.E. 2009. A minimalist two-level foundation for constructive mathematics. *Annals of Pure and Applied Logic* 160(3): 319–354.

Maietti, M.E., and S. Maschio. 2015. An extensional Kleene realizability model for the Minimalist Foundation. In *20th International Conference on Types for Proofs and Programs (TYPES 2014)*, volume 39 of *Leibniz International Proceedings in Informatics (LIPIcs)*, ed. P.L.H. Herbelin, and M. Sozeau, 162–186.

Maietti, M.E., and G. Rosolini. 2012. Elementary quotient completion. *Theory and Applications of Categories* 27(17): 463.

Maietti, M.E., and G. Sambin. 2005. Toward a minimalist foundation for constructive mathematics. In *From sets and types to topology and analysis: Practicable foundations for constructive mathematics*, Number 48 in Oxford Logic Guides, ed. L. Crosilla and P. Schuster, 91–114. Oxford University Press.

Martin-Löf, P. 1975. An intuitionistic theory of types: Predicative part. In *Logic colloquium '73 (Bristol)*, volume 80 of *Studies in logic and the foundations of mathematics*, 73–118. Amsterdam: North-Holland.

Martin-Löf, P. 1984. *Intuitionistic type theory. Notes by G. Sambin of a series of lectures given in Padua, June 1980*. Naples: Bibliopolis.

Maschio, S. 2015. On the distinction between sets and classes: A categorical perspective. In *From logic to practice*, volume 308 of *Boston studies in the philosophy and history of science*, 185–199. chapter 10: Springer.

Maschio, S. 2020. Numerical existence property and categories with an internal copy. *Logica Universalis* 14(3): 383–394.

McCarty, C. 1986. Realizability and recursive set theory. *Annals of Pure and Applied Logic* 32(2): 153–183.

Nordström, B., K. Petersson, and J.M. Smith. 1990. *Programming in Martin-Löf's Type Theory, an introduction*. Oxford University Press.

Poincaré, H. 1906. Les mathématiques et la logique. *Revue de Métaphysique et de Morale* 14(3): 294–317.

Rathjen, M. 2005. The disjunction and related properties for constructive Zermelo-Fraenkel set theory. *The Journal of Symbolic Logic* 70(4): 1233–1254.

Rosolini, G. 1982. Un modello per la teoria intuizionista degli insiemi. In *Atti degli Incontri di Logica Matematica, Siena*.

Shapiro, S. 2005. Categories, structures, and the Frege-Hilbert controversy: The status of meta-mathematics. *The Philosophy of Mathematics (3)* 13(1): 61–77.

Simpson, A.K. 1999. Elementary axioms for categories of classes (extended abstract). In *14th Symposium on Logic in Computer Science (Trento, 1999)*, 77–85. , Los Alamitos: IEEE Computer Socirty.

Swan, A.W. 2014. CZF does not have the existence property. *Annals of Pure and Applied Logic* 165(5): 1115–1147.

Troelstra, A.S. 1971. Notions of realizability for intuitionistic arithmetic and intuitionistic arithmetic in all finite types. In *Proceedings of the Second Scandinavian Logic Symposium (Oslo, 1970)*, 369–405. Studies in logic and the foundations of math., Vol. 63.

Troelstra, A.S., and D. van Dalen. 1988. Constructivism in mathematics, an introduction, vol. I. In *Studies in logic and the foundations of mathematics*. North-Holland.

Van Oosten, J. 2008. *Realizability: An introduction to its categorical side*, volume 152 of *Studies in logic and foundations of mathematics*. Elsevier.