On determinacy and arithmetic
First steps towards an account of the determinacy of arithmetic

Dopico Fernandez, Pablo

Awarding institution:
King's College London

On determinacy and arithmetic

First steps towards an account of
the determinacy of arithmetic

Pablo Dopico Fernandez

Supervised by
Dr Carlo Nicolai

A thesis presented for the degree of
Master in Philosophical Studies in Philosophy

Department of Philosophy
King's College London
September 2022

To Antonio.
And to the memory of Mariano,
for it was next to his bed that
the idea behind this project took shape

Abstract

Philosophers and mathematicians typically assume that arithmetic is determinate, i.e., that all well-formed arithmetical statements have a determinate truth-value. By contrast, many believe that set theory could be indeterminate. This thesis intends to take the first steps towards grounding the determinacy of arithmetic thus understood and establishing the contrast with the case of set theory.

The first chapter of the thesis surveys extant material on the determinacy of arithmetic. Thus, it first introduces the notion of arithmetical determinacy and presents a challenge in the literature that urges us to account for the determinacy of arithmetic. Then, it outlines and assesses some of the most influential arguments against the determinacy of arithmetic, paying special attention to Putnam's model-theoretic arguments [Put80, Put81], and the responses that different philosophers have offered against these arguments.

The second chapter takes a step back, and notices that the nature of the project requires an understanding of the broader picture regarding determinacy as a feature of formal systems. Thus, it addresses the more general phenomenon of truth-theoretic determinacy: it aims at laying the foundations for a theory of determinate truth that allows us to speak about determinacy from within the object mathematical theories themselves. To this extent, it draws on three theories that present a desirable trait for a theory of truth, namely supervaluational truth. These theories are: the van Fraassen-Kripke fixed-point semantics [Kri75], Stern's supervaluational-style truth [Ste18], and McGee's theory of definite truth [McG91]. In the chapter, special attention is given to McGee's theory, for it displays an attractive feature: material adequacy for the truth predicate.

Finally, the last chapter advances a first defence of the determinacy of arithmetic, based on what is often known as Isaacson's thesis [Isa87, Isa92]. According to this thesis, arithmetical truth coincides with provability in Peano Arithmetic; thus, if the thesis holds, arithmetical determinacy is guaranteed, for all there is to arithmetic is provability or refutability in Peano Arithmetic. The chapter advances a challenge to the thesis, according to which the latter could entail that Peano Arithmetic proves non-arithmetical truths. It then sets out to argue that the challenge can be withstood, since all seemingly non-arithmetical truths could, in fact, be shown to be arithmetical.

Acknowledgements

It is difficult to know when a work belongs to oneself. I do not think this one does to me. This is not to say that I have not done my utmost for this thesis—it just means that it would be unfair to attribute to myself a merit which is shared. Everyone I will mention here deserves, to a greater or lesser extent, whatever credit I could be given for this dissertation. Two years and thirty thousand words might not look like much, but even the smallest achievements often require a good deal of support from the people that care about you. And you showed me you care about me.

Needless to say, although I will say it nonetheless, Carlo Nicolai is to be thanked in the first place. Of giants and their shoulders I do not know much; but I do know that, in order to see further, one needs to see in the first place. And I would have been blindfolded too frequently if it were not for his guidance and patience.

También quiero dar las gracias a mi familia. A mis hermanas, por quererme, y por dejarse querer. A mis padres, que siempre me han apoyado, a pesar de no tener muy claro qué son la lógica y la filosofía de las matemáticas. Y a mis abuelos, que siempre me han apoyado, a pesar de que ninguno de los tres tuviese muy claro qué demonios hacía su nieto en la universidad.

I want to thank Lucía, because it is only by her side and her tranquility that I can eventually focus on the woosh of the trees and the odor of the berries. She has given me more love and understanding than anyone can ask for.

I want to thank my PGR fellows, for the burden is communal. Special thanks go to Anna, because we have shared good and bad. Another mention goes to Matt Soteriou, for being immensely supportive as an MPhilStud Lead. I want to thank the Logic quartet, with an outstanding mention to Beatrice Buonaguidi, who knows some of my proofs almost better than I do.

I want to thank Luca, for bearing with me for an entire academic year; and Sole, for bearing with me during the summer as well. I want to thank Sergi, Irene and Rocio, because we survived lockdowns together. I want to thank my London friends, who hopefully will forgive me for not naming each of them; and I want to thank Claudia, for being there when it all began.

Just to finish: I want to thank Luca Incurvati and Johannes Stern for their willingness to be my examiners; Daniel Isaacson for discussing my work with me so enthusiastically; and Maciej Glowacki for the valuable feedback. Previous versions of this thesis (in particular, of chapter 3) were presented at the Konstanz Summer School for the Philosophy of Mathematics and at Professor Niebergall's Logic Colloquium, as well as various departmental and inter-departmental seminars and talks; I am truly grateful to the respective audiences for all their comments.

Contents

Introduction

Consider the following two questions:

a) Is every even natural number greater than two the sum of two primes?

b) Is every projective set of reals Lebesgue-measurable?[1]

In answering these questions, even the most knowledgeable of the mathematicians will be cautious, albeit for different reasons. In the case of a), we are dealing with Goldbach's famous conjecture. The conjecture, which pertains to the mathematical field of arithmetic,[2] has neither been proved nor disproved as of today—almost 400 years after it was first presented. A quick Google search asserts that the conjecture holds for the first 4×10^{18} integers, so it is very likely that it will hold for any number that any of us can think of. Similarly, none of the greatest mathematicians in the history of the subject doubted the truth of the conjecture. And yet, the failure to find a proof after so many years has led some to believe that it could, after all, be the case that the conjecture has no proof. Given that proof is the standard of truth in mathematics, then, how are we to give a determinate answer to question a)?

In the case of b), things are slightly different. The claim that all projective sets of reals are Lebesgue-measurable, which belongs to the field of descriptive set theory, is known to be independent of **ZFC** (Zermelo-Fraenkel with the Axiom of Choice), that is, neither the claim nor its negation follow from the axioms of ZFC. As it happens, the theory **ZFC** is often seen as the best description of the set-theoretic universe, so the fact that its axioms do not settle the question of Lebesgue-measurability for projective sets also suggests—or so some argue—that this is an

[1] For the notions of 'projective set' and 'Lebesgue measurability' see [Mos80, ch.1-2].

[2] We shall use the terms 'arithmetic' and 'number theory' interchangeably. What nowadays, outside philosophical contexts, is known as 'arithmetic', we shall call 'elementary arithmetic'.

instance of indeterminacy in set theory, and mathematics more generally. In this regard, the situation is similar to that of Cantor's well-known continuum hypothesis (CH), which states that there is no set of real numbers whose cardinality lies between the size of the set of natural numbers ($\aleph_0$) and that of the set of all reals: CH is also independent of ZFC.

Having said that, there is a fundamental aspect that differentiates questions a) and b). This is no other than the fact that philosophers and mathematicians have strong intuitions to the effect that arithmetic is determinate, and hence that there is a definite answer to question a)—whether provable or not; while many of them have serious doubts as to the determinacy of set theory, and higher-order mathematics. The precise understanding of 'determinate' and 'determinacy' will be explored in chapter 1 of this thesis; for now, let us informally say that an area of mathematics is determinate if there is a determinate answer to all meaningful questions in the spirit of a) and b) that can be posed for that area. Hence, the difference between a) and b) is that most of us firmly believe that there is a determinate answer to all meaningful arithmetical questions, and yet some of us suspend judgement (or directly deny) the existence of a determinate answer to all meaningful set-theoretic questions.

Why this happens is an interesting question. José Ferreirós [Fer16, ch.7] gives one answer. He argues, quite convincingly, that our confidence in the certainty of arithmetic is due to the fact that we recognise the principles of arithmetic as true of our most elementary counting practices. Counting, in turn, is an essential technique for our interaction with, and understanding of, the world as human beings. He writes:

> The cognitive complexity of counting is certainly very high, incomparably higher than that of subitizing, but the most noteworthy aspects of this practice are its stability, reliability, learnability, and intersubjectivity. I dare to suggest the idea that mathematics, in its most elementary strata, may be the best expression of that which we humans have in common, merely in virtue of being human. [Fer16, 188]

Therefore, as a result of our deep internalization of this elementary side of mathematics, arithmetic would necessarily be seen as certain and determinate. Once again, Ferreirós' is just one possible answer; there might be others. In any case, it is not that question that we are interested in, but rather: how can we show the correctness of that conviction? That is, how can we establish that arithmetic is determinate? And how does this differ, if at all, from the case of set theory, which we believe that could be indeterminate? This thesis is a first step towards answering these worries.

As a first step, it is an incomplete piece of work, which assumes the continuation of the project for the remaining of my doctoral studies. Hence, this MPhilStud thesis will serve as the foundations of that bigger project. That is: the reader will not find here a solution to the driving research question (what makes arithmetic determinate), but they will at least find the first pieces of what I hope will become the solution.

Now, the question of what, if anything, makes arithmetic determinate has been posited several times in the contemporary literature. Some prominent responses include Hartry Field [Fie94], Leon Horsten and Volker Halbach [HH05], or Sharon Berry [Ber21]. In chapter 1, we will explore these and other accounts, in order to set the stage for the project. This first chapter is intended as a review of the literature on the determinacy of arithmetic. Thus, the first thing we do in the chapter is to explain what 'arithmetical determinacy' means in our context: we outline a semantic notion of determinacy, and introduce **arithmetical determinacy** as the thesis that all arithmetical statements have a determinate truth-value. We then present Warren and Waxman's metasemantic challenge to arithmetical determinacy [WW20a], which is of vital importance for us because it is the challenge that we want our project to eventually meet. Right after, we introduce what seems like the biggest threat to **arithmetical determinacy**, namely, Putnam's model-theoretic arguments [Put80, Put81]. We offer and critically assess different ways in which the literature has tried to counter Putnam's arguments, noting that none of them seems entirely satisfactory.

The second chapter tries to look at the overall picture, and to provide the basis for a thorough understanding of the notion of determinacy. In particular, we argue that we need of a theory of determinate truth with which to explore the idea of determinacy in the very object theory whose determinacy is being investigated. Moreover, we want that theory to display two desirable features of truth: a guarantee that logical truths come out as determinate, that is, a supervaluational truth; and material adequacy for truth and determinate truth, i.e., Tarski's T-convention for the meta-language. Thus, we trace the connections between three theories of truth: two of them, by van Fraassen-Kripke [Kri75], and by Johannes Stern [Ste18], are essentially supervaluational; the third one, by Van McGee [McG91], is supervaluational in spirit *and* materially adequate. We also characterize the theory proposed by McGee, a task McGee himself did not undertake.

Finally, the third chapter leaves aside the search for a theory of determinacy, and proposes a way to secure arithmetical determinacy that, we argue, is a partial response to our research question. In particular, we adopt what the literature has called 'Isaacson's thesis': the thesis that Peano Arithmetic is complete and sound with respect to arithmetical truth [Isa87, Isa92]. The proposal is based on an epistemological reading that Daniel Isaacson makes of the notion of arithmetical truth, according to which a statement is arithmetical when it follows from our understanding of the natural numbers alone. We explain that, if Isaacson's thesis is true, then arithmetical determinacy is secured, as all there is to arithmetic is provability and refutability in Peano Arithmetic. Then, we notice how Isaacson's thesis may lead to the conclusion that Peano Arithmetic is *inadequate*, a term we introduce to refer to the fact that the theory may prove truths that are not arithmetical in nature. If this was the case, we argue, Isaacson's thesis would lose much of its appeal. The main goal of the chapter is thus to defend the conjecture that all those seemingly non-arithmetical truths are, in fact, arithmetical by Isaacson's standards.

Chapter 1

Throughout the introduction, we have stated that the goal of this project is to find the way to justify the claim that arithmetic is determinate. In this first chapter, we offer an extensive but non-exhaustive review of the literature on the determinacy of arithmetic. The aim here is to understand the magnitude of the challenge behind the project, and to set the stage for the remaining of it.

We open with an attempt to understand what is meant by determinacy in the context of the philosophy of mathematics. We thus provide a semantic conception of it, and explain what it takes for arithmetic to be determinate on that basis. After that, we present a first challenge to the determinacy of arithmetic that, we indicate, arises mostly for formalist and conventionalist views of mathematics. The following section presents problems for the determinacy of arithmetic that also apply to the realist about mathematics, the so-called model-theoretic arguments. We then explain, for two sections, some prominent responses to these model-theoretic arguments. Finally, we introduce two alternative ways of arguing for the determinacy of arithmetic.

1.1 What is mathematical determinacy

Explaining what is meant by mathematical determinacy is already a difficult challenge. The notion of determinacy is ubiquitous in Philosophy, and is often better understood by examining what is meant when we say that something is indeterminate. As it happens, there is a plethora of opposing views as to what can ground such a claim. Let us consider a classical example: a person, say John, who has too little hair to be confident on him not being bald, but an amount of hair such that we cannot really say if he is bald either. There are various routes one can take to account for this example, but three stand out as particularly popular in the literature.

One could argue, for instance, that indeterminacy is epistemic in nature, and that instances of indeterminacy are just a manifestation of some unknowability of the subject matter in question—a deep, ineffable form of ignorance. Even if they may be a fact of the matter about whether John is bald or not, we simply cannot know it—something like this is what Timothy Williamson [Wil94] proposes. This theory, or collection of related theories, has received the name of *epistemicism*, for obvious reasons. Alternatively, one can frame indeterminacy claims in metaphysical terms: whether John is bald or not is a primitive, metaphysical fact, not to be understood via any reduction. An interesting account along these lines is developed by Barnes and Williams [BW11]. Finally, the third route to account for cases of indeterminacy or vagueness is semantic. Indeterminacy is a problem of our language, not of our knowledge nor of the world. Thus, one can argue that the semantical property of bivalence fails for certain claims of natural language involving indeterminate terms or predicates. One may then want to explain this failure further. For instance, the early Kit Fine [Fin75] took indeterminacy to be a 'deficiency of meaning' of the terms involved. Accordingly, a logic for determinacy must model the ways in which the meaning of the problematic terms can be sharpened.

What about mathematics? As before, one can believe that, even if there is a matter of fact about all mathematical claims, some are indeterminate because there is no human way to get to know their truth or falsity. That is: it will either be the case that every even number greater than two is the sum of two primes, or it will not, but such knowledge is unavailable to us. In his epistemicist line, Williamson seems to defend this account [Wil94, 204].[1] Clearly, this position seems to stand in opposition to anyone who believes the Hilbertian dictum that, in mathematics, there is no *ignorabimus*. But it is not wholly incompatible with being a die-hard mathematical Platonist that fights back any claim to the effect that the ontology of mathematics really is determinate.[2] This being said, general consensus among die-hard Platonists has it that mathematics is a determinate matter, and that the truth or falsity of a seemingly indeterminate mathematical statement can eventually be known; Kurt Gödel is perhaps the most famous example of such attitude [Gö90].[3]

[1] Although, admittedly, Williamson does not say that Goldbach's conjecture is an example of an *indeterminate* mathematical claim—only that it is an example of a necessary but unknowable truth.

[2] By die-hard mathematical Platonist, I mean a Platonist in the traditional sense: a supporter of the theses that (i) there are mathematical objects, (ii) these objects are abstract, and (iii) they are independent of intelligent beings in any possible way (see [Lin18]). I also exclude any form of pluralist or plenitudious mathematical Platonism.

[3] Interestingly enough, Gödel argued that his anti-epistemicist convictions do not derive from his belief in Platonism; rather, they are the result of the fact that we undeniably possess an intuition for mathematics that plays in the mathematical epistemic enterprise a role similar to perception in physics. Even if there were no mathematical objects, '[t]he mere psychological fact of the existence of an intuition

In contrast to the case of epistemicism, and as far as I can tell, the possible indeterminacy of mathematics has never been considered in strictly metaphysical terms, that is, it has never been explained as a brute metaphysical fact. Moreover, it is not clear how to flesh this idea out. One might argue that, under the so-called full-blooded or plenitudious Platonism, the mathematical claim '$2^{\aleph_0} = \aleph_1$', which is just the continuum hypothesis, is indeterminate due to metaphysical reasons. After all, plenitudious Platonism—first introduced by Mark Balaguer [Bal98]—holds that all consistent mathematical theories truly describe a part of an objective, abstract, and agent-independent mathematical realm;[4] and so '$2^{\aleph_0} = \aleph_1$' holds in the part of the realm described by **ZFC** + CH (or **ZFC** + $V = L$, for that matter), and '$2^{\aleph_0} \neq \aleph_1$' holds in the part of the realm described by **ZFC** + ¬CH. Hence, the argument would go, truth-value of '$2^{\aleph_0} = \aleph_1$' is indeterminate. But that is the point: even if there is a metaphysical explanation behind, indeterminacy arises at the sentential level. It is the claim '$2^{\aleph_0} = \aleph_1$' that fails to have a determinate referent. The situation can be somehow analogous to what happens in uttering 'Alice's dog has a black spot on the back', while in fact Alice has two dogs, Toby and Giorgio, and only one of them has a black spot on the back. There is no metaphysical indeterminacy, but an indeterminacy of reference. If we were to index the sets of the different parts of the mathematical realm as described by the mutually incompatible theories—just like one specifies whether they refer to Toby or Giorgio—indeterminacy disappears. Thus, '$[2^{\aleph_0} = \aleph_1]_\alpha$', where $[\cdot]_\alpha$ serves to single out the mathematical realm as described by **ZFC** + $V = L$, is not indeterminate anymore, but true. In sum, we are dealing with a semantic conception of indeterminacy.

This is a symptom of a more general fact, namely, that, rare exceptions aside (including Williamson), the favoured view of indeterminacy in mathematics has been essentially semantic. In this work, we adhere to the tradition, hence considering determinacy and indeterminacy to be (possibly) tied to mathematical statements themselves, and not to some supposed underlying ontology or to what we can know about mathematics. Thus, a first, more precise approximation to mathematical (in)determinacy thus understood has been made explicit by Warren and Waxman [WW20a, 478], and runs as follows:

> **Mathematical indeterminacy**: the thesis that some mathematical statements do not have a determinate truth-value, i.e., that some math-

... suffices to give meaning to the question of the truth or falsity of proposition like Cantor's continuum hypothesis', as well as Goldbach's conjecture [Gö90, 268-9].

[4]The 'consistent' requirement can be relaxed to give rise to what J.C. Beall calls 'really' full blooded Platonism [Bea99].

ematical statements are neither determinately true nor determinately false.[5]

Accordingly, **mathematical determinacy** is the thesis that all mathematical statements have a determinate truth-value.

Of course, this has to be nuanced in at least the two following ways. First, a mathematical statement can only count as such if it is a *well-formed* mathematical statement. Clearly, that will depend on the particular language we are dealing with, but it seems evident that many arbitrary combinations of mathematical symbols will not count as well-formed statements. In these cases, we shall not expect them to have a truth-value, and this should not undermine our confidence in mathematical determinacy. Secondly, in a similar vein, if a seemingly mathematical statement includes a predicate that may count as vague or fuzzy outside mathematical contexts, the indeterminacy may permeate the mathematical statement itself. A good example, adapted from [War20, 204], is when we say: '10 is a small number'. That people may rightfully disagree on the truth-value of this statement is no threat to the determinacy of mathematics, for we shall not count it as a genuine mathematical statement—however one wants to understand 'genuine'.[6]

Two further remarks are in place here: unlike the case of epistemicism, Warren and Waxman defend that **mathematical indeterminacy** implies that there is no fact of the matter with respect to the truth or falsity of the statement or statements that have no determinate truth-value. Furthermore, they also contrast their definition of indeterminacy with that of pluralism. For them, pluralism for a subject-matter D is the idea that there are multiple, (seemingly) incompatible theories for D that are equally correct. As in the case of plenitudious Platonism (which can be used to support a form of pluralism), when a mathematical statement is uttered without the specification of a particular theoretical framework, mathematical pluralism arguably entails mathematical indeterminacy: for two incompatible but accepted theories T_1 and T_2, there will be no determinate truth-value for the sentences for which T_1 and T_2 differ. However, the reverse need not be the case.

[5]We use the term 'statement' instead of Warren and Waxman's term 'claim'.

[6]A notable exception is the predicate 'being a set', or the notion of set more generally. Even though the term plays a central role in mathematics, mathematicians and philosophers alike have often found themselves quarrelling over differing conceptions of set—see [Inc20, Ch.1]. For many, this issue lies at the heart of the indeterminacy of set theory, and perhaps of the indeterminacy of mathematics more generally (for an alternative view see e.g. [Mar01]).

1.2 The first challenge to arithmetical determinacy

The view that **mathematical indeterminacy** is true is not extremely implausible, and it has received substantial support from well-known figures.[7] Those who adhere to it often point at the independence of the Continuum Hypothesis and other set-theoretic claims as evidence that the concept of set—or related notions, such as that of the cumulative hierarchy—is vague or indeterminate (and, perhaps, under-determined), with the implication that certain statements lack a determinate truth-value. But consider **arithmetical indeterminacy**, the thesis that is obtained by replacing 'mathematical' for 'arithmetical' all along in our definition of **mathematical indeterminacy**. Now, **arithmetical indeterminacy** is a very unpopular thesis, even among those who argue that set theory is indeterminate (cf. [FFMS00, 410]). There seems to be a fact of the matter about *every* arithmetical statement; the structure of the natural numbers and, by extension, the standard model of arithmetic, seem clearly determinate in all regards.

Now, Warren and Waxman [WW20a] have produced an argument which aims to deliver the conclusion that arithmetic is indeterminate. The argument unfolds in various steps. First, they say, a rather intuitive argument in favour of the indeterminacy of CH—and, consequently, of set theory—runs as follows:

1. CH is independent of **ZFC**.
2. A sentence of the language of set theory is indeterminate iff it is independent of **ZFC**.

∴ CH is indeterminate.

But then, if we let R be the Rosser sentence for **PA**,[8] we can introduce the following argument:

1. R is independent of **PA**.
2. A sentence of the language of arithmetic is indeterminate iff it is independent of **PA**.

∴ R is indeterminate.

[7] See e.g. [FFMS00, 402] and [Mos67].

[8] The Rosser sentence is a sentence that says of itself that, if there exists a proof of it in **PA**, then there exists a smaller proof in **PA** of its negation. We could have equally introduced the Gödel sentence for **PA**.

What Waxman and Warren claim is that the two arguments are strikingly similar, and that we may need to take the second argument seriously: premise 1 of both arguments is undeniable, and premise 2 captures analogous ways of understanding set-theoretic truth and arithmetical truth as provability in (or, at least, extensionally equivalent to provability in) some privileged mathematical theory. So, the argument goes, if we are to resist **arithmetical indeterminacy**, then we need to explain how the determinacy of arithmetic can arise—what they call the metasemantic challenge. Unless such a challenge is met, the argument just sketched gives us a reason to believe that arithmetic is determinate.

But does it really? There are some problems here. Warren and Waxman already note that there are relevant differences between R and CH. We have good reasons to believe that R is true (e.g., that it holds in the standard model of arithmetic), which does not happen in the case of CH. I would argue, however, that there is a more fundamental justification to reject the force of the argument, namely that the reasoning for the indeterminacy of CH is highly contestable. In particular, premise 2 is, for the most part, false. A clear example is the aforementioned statement 'All projective sets of reals are Lebesgue-measurable'. With rare exceptions (including Feferman and, perhaps, radical proponents of set-theoretic pluralism), most set theorists wholeheartedly believe that the statement is determinately true: it captures a nice, desirable, and very plausible property of sets of reals; and, while independent of **ZFC**, it can be shown to follow from rather weak large cardinal assumptions. In fact, many set theorists take it to be important evidence that at least some of these large cardinal axioms are to be accepted too as part of our description of the set-theoretic universe. The situation is very distant from that of CH, where it is known that large cardinal axioms cannot settle the hypothesis, and that axioms that do settle it are somehow much less intuitive. In sum: premise 2 can be argued to fail, according to how many set theorists and philosophers view the subject.

Now, this criticism should not lead to the conclusion that the metasemantic challenge has been dispelled. It is still necessary to explain why mathematics is determinate, rather than just assume it is; determinacy is too important of a property of our mathematical talk and practice to overlook it. In a sense, it is similar to Benacerraf's two challenges [Ben73]: it is now generally agreed that an acceptable philosophy of mathematics must be able to tell us how it is possible that our mathematical beliefs are reliable—put otherwise, how it is possible that we have mathematical knowledge at all. Most relevantly for our topic of investigation, it is still necessary to explain why *arithmetic* is determinate; we want to be able to say that the claims 'All numbers greater than 2 is the sum of two primes' and '$2+2=4$'

have a determinate truth-value.

In fact, and as the Introduction to this thesis made clear, the goal of this project consists precisely in meeting the metasemantic challenge. We want to explain how the determinacy of arithmetic arises, and how it stands in opposition to the (possible) indeterminacy of set theory. After all, arithmetic, unlike much of higher-order mathematics, is part of our everyday understanding of, and interaction with, the world around us. We use counting and elementary arithmetical operations all the time. Science, engineering, and all sorts of human artifacts rely on maths and measurements, and ultimately on those basic arithmetical operations. It would clearly seem fatal if we discovered that certain questions of such a fundamental part of our knowledge do not have a determinate answer, that is, if certain statements did not have a determinate truth-value.

Something to notice, though, is that the challenge is not a pressing one for those who hold certain views on the philosophy of mathematics. For instance, as acknowledged by Warren [War20, 204], the conventionalist about mathematics needs to address the challenge, on pain of having to admit that our conventions are plainly faulty. This possibly applies to the formalist about mathematics as well. On the contrary, a fictionalist about mathematics who believes that no mathematical statement is true would not feel compelled to meet the challenge.

Likewise, and as Warren [War20, 205] rightly points out, the die-hard Platonist can appeal to 'theory-transcendent' facts in order to settle the determinacy of, for instance, arithmetic. The idea, using Warren's terms, would be that the ontological facts 'outstrip' our practice. Our theory just presents limitations when it comes to accounting for certain worldly facts—but the facts are still there. Of course, this solution is only available to someone who can adequately argue in favour of die-hard Platonism—and that is no small task. Moreover, it is unclear to me whether this proposal is not actually collapsing into a form of epistemicism. If the idea is that the facts will *always* outstrip the practice, there is a gloss of unknowability in the Platonist's appeal. Anyway, this matters little for Warren and Waxman, for appealing to metaphysics in order to explain determinacy is ruled out by their imposition of what they call the 'metaphysical constraint'. The constraint, which arises out of naturalistic concerns, basically states that abstract objects should play no role in mathematical metasemantics. That is, we cannot point at some supposed well-defined and determined abstract, mathematical realm to answer questions about meaning and reference. Together with the so-called cognitive constraint (i.e., the idea that we have no non-computational mental powers), which blocks an appeal to non-recursive theories that do happen to be complete, we are again in dire need for an account of mathematical (or, at least, arithmetical) determinacy.

1.3 Putnam's model-theoretic argument(s)

We just saw how, under the metaphysical constraint, the metasemantic challenge raised by Warren and Waxman remains. But one may have independent reasons to resist the metasemantic constraint. Is then the Platonist on safe grounds as regards the determinacy of, at least, arithmetic?

There are strong reasons to doubt it. For, on the first place, the phantom of full-blooded Platonism (FBP), which we already mentioned, awakes as soon as we adopt a Platonist perspective. After all, FBP seems to fare better than die-hard Platonism in many respects. For instance, it (arguably) meets Benacerraf's challenge, and is able to account for the way in which mathematicians keenly study all sorts of theories.[9] But, as we saw, the full-blooded platonist is happy to assert that '$2+2=4$' has no determinate truth-value, contradicting thus the kind of determinacy we are after.

All in all, the biggest challenge to securing determinacy by adopting a form of traditional Platonism lies somewhere else, namely in Hilary Putnam's well-known model-theoretic argument against realism. Or, rather, in his model-theoretic argument*s*, for there are three of them.[10] We shall begin by briefly sketching each of them.

1.3.1 The arguments

The first argument Putnam presents in his paper *Models and reality* [Put80] is sometimes known as the argument from Löwenheim-Skolem, for it uses this well-known model-theoretic result. The idea of the argument is simple. Suppose that your preferred theory T 'says' that it has an *uncountable* set S. Then, by the Löwenheim-Skolem theorem, T has a *countable* model—one where the statement 'S is uncountable' still holds. As it is well-known, the contradiction is only apparent; the way models see themselves does not reflect how they are 'in reality'. This is why Skolem would say that 'countable' and similar notions are relative to a model. The problem for Putnam is that one cannot rule out the possibility of living in a non-standard model of the sort just outlined, where things do not mean what they should 'in reality'. Any theoretical attempt to single out the standard model as the universe of

[9] These and other points are found in [Bal98, Ch.3].

[10] I am grateful to Tim Button for clarifying this point, as much of the literature does not distinguish the arguments.

our discourse must be done within a theoretical framework, formalisable within some theory T' an unintended model of which we cannot rule out. That is why this attempt to neutralise Putnam's argument has often been known as the 'just-more-theory' manoeuvre. Moreover, the challenge does not end with the notion of countability; one could argue that the same applies to all sorts of theoretical terms, from our everyday vocabulary to the sharpest scientific jargon. What's key here is our incapacity to pick out the extensions of these terms correctly due to the existence of non-standard models that we cannot leave aside on the basis of theory alone. As a result, what obtains is referential indeterminacy, which amounts to a form of semantic indeterminacy.

What about *observational* constraints? Can't our experience of the world rule out non-standard models for these terms? Not for Putnam: we would fall prey to a similar worry. This being said, very few people have conceded Putnam this last point and, at least when it comes to our everyday terminology, it is generally believed that our operational constraints can and indeed do fix the referents of our everyday language (see e.g. [Fie94, §2]; [BW16, 286-7]). But since we are here interested in *mathematical* determinacy alone, it seems that our observational constraints will certainly not fix the referent of, say, the term 'the class of all projective sets of reals'.

Let's check the second model-theoretic argument. Button and Walsh [BW16, BW18] have baptised this argument the 'push-through construction', but it has also appear in the literature under the name of 'permutation argument' (see [HW17, Gas11, But11]). The argument is based on the idea that, given a model for a theory T, we can find many isomorphic copies of it for which the domain has been permuted. Putnam explains it in plain terms in the following way. Suppose that we have a model over a language that contains the terms 'cat' and 'cherry', and whose domain quantifies over cats, cherries, and possibly other things. Then

> if the number of cats happens to be equal to the number of cherries, then it follows from theorems in the theory of models ... that there is a reinterpretation of the entire language that leaves all sentences unchanged in truth value while permuting the extensions of 'cat' and 'cherry'. By the techniques just mentioned, such reinterpretations can be constructed so as to preserve all operational and theoretical constraints. [Put81, 41]

If so, how could we ever be sure that the referent of our term 'cat' is what we want it to be? Well, once again, we can downplay Putnam's conclusion: observational constraints should suffice here. We are very much acquainted with cats and cherries. But what happens when the discourse is mathematical? The very same argument can be made of, say, a model of the natural numbers that permutes each even number n for the number $n + 1$—so that 0 and 1 swap places, 2 and 3 swap places, and so on.

Finally, the third model-theoretic argument may be known as the constructivisation argument, and it is more involved than the previous two. As Button [But11, 323] explains, this is an argument directed at the realist about sets. A realist about sets in the standard fashion would have it that there is a fact of the matter about whether Gödel's Axiom of Constructibility, $V = L$, holds or not. The axiom basically states that all sets are constructible, in a technical sense of the word (for more, see e.g. [Kun13, §II.6]). As Putnam [Put80, 467] notes, Gödel believed it to be false, and so do most set-theorists. How can we establish that conclusion? According to Putnam, it cannot be through theoretical constraints. On his view, a model meets the theoretical constraints for a set theory as soon as our preferred axioms are satisfied in the model. Since these would be the axioms of **ZF** (or perhaps **ZFC**), and $V = L$ is consistent with these axioms (i.e., there are models of $V = L$ that also satisfy **ZF**), the theoretical constraints will not determine that the Axiom of Constructibility is false. As a result, it must be our operational constraints that falsify the axiom. This would amount to something like us coming across a non-constructible real number: since the Axiom of Constructibility asserts that the set of all constructible sets (L) is equal to the whole set-theoretic universe (V), an empirical encounter with—perhaps a measurement that yields—a real number (i.e., a set) that we knew to be non-constructible would show that $V \neq L$. But then, Putnam proves the following result:

Theorem 1 (Putnam)**.** *For any countable set of real numbers S, there is an ω-model $\mathcal{M}$ such that $\mathcal{M} \models$ **ZF** $+ V = L$ and S is represented in $\mathcal{M}$.*

That S is 'represented in $\mathcal{M}$' here just means that S can be coded up as a single real number s (there are standard techniques to do that), and that $\mathcal{M}$ contains the single real s—maybe as a Dedekind cut, or a Cauchy sequence (see e.g. [Kun12, §I.15]). All in all, the conclusion is that, whatever non-constructible 'in reality' real number s could exist, coming across this number would not settle the question of whether $V = L$ or not, for we could have encountered this real number in a constructible universe too. But this seems the only chance we could have of, observationally, decide the truth of the Axiom of Constructibility. So the realist about sets cannot be right, unless they posit that we are in possession of some super-natural ability to grasp the truth of set-theoretic statements—which is a poor epistemological move. Thus, Putnam concludes, 'the "relativity of set-theoretic notions" extends to a relativity of the truth value of "$V = L$" ' [Put80, 469]. That is, not only notions as 'countability' or 'set' are relative, the truth-value of general set-theoretic statements also is—another ghost of indeterminacy![11]

1.3.2 Putnam's arguments and arithmetical determinacy

In this subsection, we are going to explore the extent to which Putnam's argument might lead to **arithmetical indeterminacy**, a thesis we described previously.

According to Timothy Bays [Bay01], it is not possible that Putnam's arguments lead to **arithmetical indeterminacy**; in fact, it is not possible for them to lead to any indeterminacy at all by themselves, for the premises on which they rest are wrong. What Bays questions, in particular, is: (1) the mathematics behind Putnam's argument, and (2) the claim that only theoretical or operational constraints would fix the reference of our terms. We shall here examine objection (1), for it is the one that has left a bigger imprint on the literature. After all, if the mathematics of the arguments are wrong, then the whole arguments, being based on model-theoretic reasoning, undeniably are too.

The mathematical objection to Putnam's argument (or, at least, to the third model-theoretic argument) can be raised by asking the question: what theory are we using to prove Theorem 1? If the answer is that we are using **ZFC**, then something must have gone terribly bad, for any model of $\mathbf{ZF} + V = L$ satisfies the Axiom of Choice, and so Theorem 1 has proved the existence of a model of **ZFC** from within **ZFC** itself. But this contradicts Gödel's second incompleteness theorem! Bays notes exactly where the proof has gone wrong. In order to prove his theorem, Putnam needs a countable model that contains the set S coded as a single real s. The inner model L will give him a model that contains s, and so he applies the downward Löwenheim-Skolem Theorem to obtain a countable elementary substructure of L that contains s.[12] But that is the problem: L, being equal to V, is a class, not a set, and the downward Löwenheim-Skolem theorem can only be safely applied to sets. So, as it stands, the argument breaks. In spite of this, the alternative, namely to carry out the proof in a stronger theory, is not better. Suppose that we proved

[11] One could discuss whether this argument was really needed. It seems to me that the original Löwenheim-Skolem argument also showed that the notion of constructibility is relative. But, if so, the very truth-value of the Axiom of Constructibility is, for it just asserts that all sets are constructible.

[12] Let $\mathcal{A}$ and $\mathcal{B}$ be models of a language $\mathcal{L}$, and $|\mathcal{A}| \subseteq |\mathcal{B}|$. $\mathcal{A}$ is an elementary substructure of $\mathcal{B}$ when, for every formula φ in $\mathcal{L}$, $\mathcal{A} \models \varphi[\sigma] \Leftrightarrow \mathcal{B} \models \varphi[\sigma]$ for every every assignment σ for φ in $|\mathcal{A}|$ [Kun13, §I.15].

the theory in **ZFC** + 'there is an inaccessible cardinal κ'. Then, we can apply the Löwenheim-Skolem theorem to obtain a countable model that represents S and satisfies **ZF** + $V = L$. But then this model will not meet out theoretical constraints in the sense given by Putnam. For one thing, it will not satisfy **ZFC**+'there is an inaccesible cardinal κ', and we surely need to say that the large cardinal axiom is part of our theoretical constraints if we are going to reason with it.

In sum: Theorem 1—and with it the constructivisation argument—seems deeply flawed. But was the argument a problem for the determinacy of arithmetic at any point? Possibly not. After all, it was an argument that aimed to show the relativity of contentious set-theoretic claims ($V = L$, possibly CH, or the Axiom of Choice), and so the indeterminacy of set theory, but we struggle to find a way in which it can lead to the conclusion that arithmetic is indeterminate.

Something different occurs in the case of the first and the second arguments. Consider the argument from Löwenheim-Skolem. Some theorists, most prominently Hartry Field, have argued that the argument applies to set theory as much as it applies to arithmetic (see e.g. [Fie94, Fie98, Fie01]). Field's argument is as follows. We can consider a quantifier $\exists_{Fin}$ for expressing that there is finitely many objects that satisfy a certain formula. This quantifier can be defined in second-order logic, or set-theoretically; or it can be taken as primitive—defined, according to Field, by the rules that govern its usage. In either case, however, this quantifier will also have standard and non-standard interpretations, i.e., number-theoretic models where formulae with the quantifier are indeed satisfied by a finite number of objects of the domain, and models where formulae with the quantifier are satisfied by an infinite number of objects—including an infinity of non-standard numbers. Now, that these models exist is also guaranteed by the appropriate mathematics.[13] Therefore, it seems that the notion of 'finitely many' is also relative. But if this notion is indeed relative, the very concept of natural number also is, for a natural number can be characterised as having finitely many predecessors.

Moreover, Bays' critique does not arise in the case of the argument from Löwenheim-Skolem, according to Button [But11]. This is so because, in Button's view, given a theory T, when no $V = L$ or any other contentious statement is involved, all Putnam's argument establishes is the conditional claim 'If T has any models, then T has unintended models that we cannot tell apart'. Now, note that, by the Completeness Theorem, the antecedent of this conditional is equivalent to 'T is consistent'. And, unless one is a fully committed fictionalist about mathematics,

[13] But note that here we make use of the compactness theorem, not the Löwenheim-Skolem [Fie94, 414].

one will accept that our preferred theories, say **PA**, are true. However, Button continues, truth entails consistency.[14] Therefore, the realist must concede Putnam the antecedent he needs to conclude that our preferred theory has unintended models that we cannot tell apart.

If Button is right, then it seems that the argument Field provides should also be correct, and Putnam's argument threatens the determinacy of arithmetic. However, I believe that there is room to disagree with Button. We are now witnessing a renovated interest in the literature around the notion of *implicit commitment*, which aims to capture the idea that there may be certain components (especially, formal statements) implicit in the acceptance of a theory that are not available in the theory, or even formalizable in its corresponding language (see [NP19]). The lively debate shows that it is not entirely clear whether accepting a theory also implies accepting properties of it such as soundness or consistency, especially in light of the so-called foundational approaches to theories (e.g., Tait's link between finitism and **PRA**). Therefore, we should be wary of Button's claim that accepting T entails that one is committed to $\mathrm{Con}(T)$, and even less to the truth of it, i.e., a reflection principle. In any case, we will not pursue this objection further, and will continue as if indeed Putnam's argument from Löwenheim-Skolem escaped Bays' criticism, and as if it indeed threatened the determinacy of arithmetic, as Field wants us to believe.

Finally, let's consider the extent to which the second argument leads towards **arithmetical indeterminacy**. The push-through or permutation argument as applied to arithmetic seems to naturally lead to the idea that our preferred arithmetical theory singles out a model only up to isomorphism. There are infinitely many isomorphic copies of the standard model of arithmetic, the argument will go, and how are we to pick one? Maybe the model we live in, even if isomorphic to the model we would like to single out, is such that the term '302' takes the third position, and the term '3' takes the hundred and fifty third position. One may be then tempted to conclude that our arithmetical terms do not refer determinately, but that *the theory as a whole does* [BW16, 287]. After all, it singles out the appropriate sort of structure, namely that of the natural numbers, in the form of an *isomorphism type*.[15] Does this bring any comfort? Maybe not to the traditional Platonist who wants each and every term to refer to the 'right' number of the natural

[14] The idea would be something like: T is true is akin to the uniform reflection principle for T, and T plus the uniform reflection principle already implies $\mathrm{Con}(T)$.

[15] An isomorphism type is defined as the equivalence class originated by some isomorphism, so that two structures belong to the same isomorphism type iff they are isomorphic.

number structure, but it may initially do for a structuralist about mathematics—especially for an *non-eliminative* structuralism *à la* Shapiro [Sha97] and, perhaps, Parsons [Par08].[16] Nonetheless, even then does the phantom of Putnam's argument reappear. For how are we to tell which isomorphism type does our theory selects? After all, first-order theories such as **PA** are not categorical, i.e., they do not single out a unique isomorphism type. But if we are to single out that type ourselves, then we would be producing a structure that would be subject again to the possible permutation of isomorphism types induced by Putnam's argument. That is, if we think of the different isomorphism types generated by a first-order theory, say **PA**, as a structure, then there will be structures isomorphic to it where the domain (the isomorphism types) has been permuted. In a sense, it is the just-more-theory manoeuvre again![17]

In closing this section, let's quickly recap what has been said so far. We have defined **arithmetical determinacy** as the *semantic* thesis that all arithmetical statements have a determinate truth-value. We have presented Warren and Waxman's metasemantic challenge to **arithmetical determinacy**, namely, the need to explain how arithmetical determinacy can arise in the first place. It then seemed that, if we accepted their metaphysical and cognitive constraints, the challenge becomes a difficult one for the defender of **arithmetical determinacy**. Thus, one option available to the Platonist was to reject the metaphysical constraint in the first place and ground the determinacy of arithmetic on mathematical ontology itself. However, we showed how at least two of Putnam's model-theoretic arguments (the argument from Löwenheim-Skolem and the push-through argument) put **arithmetical determinacy** at risk even for the Platonist, insofar as they amount to referential indeterminacy, and the notion of determinacy we are interested in is semantic.

The next two sections will thus be devoted to exploring two different lines of response to Putnam's threat. The first such line is directed at the argument from Löwenheim-Skolem, and the second line addresses the push-through construction.

[16]Structuralism is, roughly speaking, the view that mathematics is about structures. Non-eliminative structuralism holds that these structures are indeed abstract objects in their own right.

[17]One could even combine the permutation and the push-through argument and claim that, if we are to determine which isomorphism type we pick up, we need to ascend to some theory that is subject to the relativisation of terms induced by the Löwenheim-Skolem argument. Thus, since terms like 'isomorphism' may be relative, how can we even speak of singling out an isomorphism type? To the best of my knowledge, no one has tried to pursue this strategy, which seems to me an enhanced approach to the model-theoretic arguments.

1.4 Responding to the argument from Löwenheim-Skolem

By now, we know that Putnam's model-theoretic argument based on the Löwenheim-Skolem Theorem proceeds by noticing that whatever the theory we want to work in, we cannot rule out non-standard interpretations of the terms associated with it. If we want to fix the standard interpretations of those terms, then we would have to move to a theoretical framework that is subject to the same worry. In fact, Putnam even defends that we run into this problem as we try to fix the referents of our terms observationally. Nonetheless, as we said, this last point has been ruled out repeatedly in the literature. The key issue, however, is what happens with those terms the referents of which are not the kind of thing we are acquainted with observationally. This is, precisely, the case of mathematics.

Here's an example: it seems that the referent of the term 'the set of all projective sets of reals' is not something we will encounter out there, it is not something we will be familiar with by observational means. Accordingly, all we have to fix the referent of the term must be theoretical in nature—and this, by Putnam's argument from Löwenheim-Skolem, leads to indeterminacy.

In a series of papers, Hartry Field has explored how this affects arithmetic. As we saw above, he focuses on the notion of 'finitely many', which seems essential to understand that of natural number, and argues that the existence of non-standard arithmetical models for a quantifier that captures this notion might render the latter indeterminate. Now, Field's proposed way out consists is noticing that, while set theory is indeed the kind of mathematical field for which observational constraints will help little in fixing reference, this need not be the case with arithmetic. He thus writes:

> In my view, the problem of finding facts about usage that rule out [non-standard arithmetical] models is completely insoluble if we look only to the uses of [the notions of finiteness, ω-sequence, etc.] within pure mathematics. However, it is also my view that the problem *can* be solved by consideration of the uses of these notions in connection with the physical world. [Fie98, 259, italics in original]

Field's proposal is then the following: given that key arithmetical notions such as 'finiteness' or 'finitely many' are also used in connection with the observable reality, we shall try to use the physical world to rule out non-standard interpretations of these notions. In order to do so, of course, Field needs to assume that notions about observables are by large determinate, so that our words 'cat', 'atom' and 'red' do not pick up, respectively, cherries, pebbles, and plainly green items. But, since we no longer buy Putnam's claim that his argument from Löwenheim-Skolem also affects operational constraints, Field's assumption is somehow granted.

Now, the meat of Field's argument rests on the actual existence of an ω-sequence in the physical world. As a clarification, an ω-sequence is a sequence of elements related in the order induced by ω—a natural number-like sequence with the order as it should be 'in reality'. Now, suppose there is such an ω-sequence in reality generated by some physical process or phenomenon (we shall later give a couple of examples of how it may look like), and the elements (rather, the *urelements*) of this sequence form the set X. Then, one can define a predicate $F(Y)$ that holds for all and only the sets Y such that Y is a proper initial segment of X. Afterwards, one can define 'Z is finite' as there being a one-to-one correspondence between the elements of Z and some set Y' such that $F(Y')$ holds, and this should suffice to establish the meaning of 'finite' (and 'finitely many', 'natural number', and so on). Of course, this by itself does not rule out the existence of non-standard models where $F(\cdot)$ holds of a set X that is actually infinite. But, Field claims, it does rule out the possibility that we are living in such a non-standard model. For, being the set X composed of urelements of some type A, if we indeed lived in such a model, we would be capable of seeing that either some of the elements of X are not in fact urelements of type A in reality, or that the physical relation holding between the urelements is not the one that corresponds in reality.

Let's give an example that Field uses. Suppose that time is infinite and Archimedean in nature—in other words, that there will be an infinity of instants in time, and that any bounded time-interval is finite. Field calls these the *cosmological* assumptions [Fie94, 416]. Then, you can start your temporal ω-sequence: take the present as your departure point, and consider a sequence of events such that each two of them will bound a time interval. What you get is indeed an ω-sequence whose urelements are events, ordered by the 'earlier than' relation. Now, one defines the predicate $F(X)$ iff X is an initial segment of this infinite sequence of events, and ties the notion of finiteness to that of one-to-one corresponde between a certain set Y and a set Z such that $F(Z)$. Once again, this will suffice for us to check and be sure that we do not live in a non-standard model, since

> any such model must assign a nonstandard extension to the formula $[F(\cdot)]$; and in particular, it must either contain things that satisfy 'event'

which are not events, or it must contain pairs of events which satisfy 'earlier than' or 'at least one second apart' even though the first is not earlier than the second or the two are not one second apart. [Fie94, 417]

As a result of this, we *can* tell apart unintended models for the notions of 'finite', 'natural number' and so on, which means that the threat to **arithmetical determinacy** that originated in the argument from Löwenheim-Skolem is dispelled.

To be sure, Field's response to Putnam is not without its objectors, beginning with Field himself. For the most evident criticism is something he already notes: that this response relies entirely on the existence of some physical or physical-like ω-sequence. It need not be a temporal sequence: it may be a spatial sequence, a sequence of entities, or something else. What's more: we do not even need to know that such a sequence exists (we would still be able to tell events and non-events apart, even if we were ignorant of their forming an ω-sequence!), but it has to exist. Otherwise, the argument collapses. Hence, it is, at best, an argument contingent on how the world could be like. And, perhaps, this makes it rather unappealing.

A second objection comes from Parsons [Par08, §49]. According to him, Field is not really entitled to claim that we would be able to recognise 'non-standard events' if we lived in a non-standard interpretation. After all, the concept of event as understood by Field (some sort of process taking place in a *finite* amount of time) is deeply mathematically charged. Therefore, Parsons considers it unreasonable to expect that, in whatever way we come to form our concept of 'event', this is considerably sharper than our number-theoretic concepts.

Now, as per the first objection, Sharon Berry ([Ber21, Berng]), has offered a version of Field's response that claims to rely only on the *possible* existence of an ω-sequence, and not in the actual existence of such a sequence. We merely sketch the solution here. She first presents a conjunction of sentences that are meant to hold with physical necessity. In the case of [Ber21], this consists of counting and induction claims concerning a coin toss (e.g. 'an object x is the 0th coin flip iff it is a coin flip and all other coin flips happen after x', or 'if the 0th coin-flip lands heads, and if for every n-th coin-flip that lands heads, the n-th+1 coin-flip also lands heads; then all coin-flips indexed by a natural number also land heads'). It also includes the axioms of Peano Arithmetic without the induction scheme. Then, Berry's assumption has it that the existence of an infinite sequence of strictly random coin flips is at least possible. It follows then that there is a possible world where there is a strictly random infinite sequence of coin tosses and the first ω coin tosses land heads. If we lived in some non-standard interpretation of the kind adduced by the argument from Löwenheim-Skolem, then the induction claim as applied to heads must be false. After all, the claim—just like Field's case—involves the use of physical vocabulary that should be rather determinate for us. But then it is clear that the claim would not hold of physical necessity, for there is a possible world where it does not hold. Hence, the possibility of living in a non-standard interpretation can be ruled out.

Berry's argument, being a tweak of Field's argument, seems much more convincing, insofar as it does not fell prey to the contingency of the latter. Moreover, it may also be the case that it avoids Parsons' objection, since the physical phenomena involved (a coin toss) is way less mathematically laden, if at all. Nonetheless, a further, important objection has been raised against Field's account, and I take it to be equally applicable to Berry's. Thus, Otavio Bueno [Bue05] has criticised Field because, Bueno argues, he holds two antagonical claims. On the one hand, Field [Fie94, §2] takes second-order quantification to be indeterminate; the exact reason why will be presented in the next section, so let's just take it for granted here. At the same time, he argues for the determinacy of the notion of finiteness. But, Otavio argues, the notion of finiteness can only be *properly* characterised via second-order logic. The reasoning is analogous to the one that generated the argument from Löwenheim-Skolem: first-order theories have models for which the notion of 'finite' is satisfied by infinite sets. However, this issue does not arise in the case of second-order theories. If we concede this point, then Field's claims are seemingly incompatible: the determinacy of the characterised notion (finiteness) is contaminated by the indeterminacy of the characterising notion (second-order quantification). Moreover, as mentioned, this seems a problem not only for Field, but also for Berry. While, admittedly, she never defends that second-order quantification is indeterminate, she is now urged to explain either how second-order quantification can be determinate, or how it could be possible for the latter to be indeterminate while our number-theoretic vocabulary remains determinate.

1.5 Responding to the push-through argument

The push-through or permutation argument leading to referential indeterminacy had two steps. First, we apply the argument once to conclude that we would have no way to tell whether our first-order theory of arithmetic picks the natural number structure in which terms such as '3' or '302' occupy the third and the three hundred and second position, and that, at best, our theory will only pick *the* structure of the natural numbers—an isomorphism type. Then, we apply the argument again to realise that our theory may not single out *the one* isomorphism type we need, since permutation of isomorphism types is also possible. Thus, we are left with radical indeterminacy of terms *and* of structure.

Fortunately, there is a *prima facie* straightforward answer to this challenge. The impossibility of singling out an isomorphism type is certainly a feature of first-order theories, but that changes with some theories formulated in second-order logic. That is, some second-order axiomatizations are *categorical*, i.e., they pick out an isomorphism type—or, equivalently, any two models of such theory are isomorphic. These include: the second-order axiomatization of Peano Arithmetic ($\mathbf{PA}^2$)—a classical result by Dedekind [Ded63]—and, for instance, the second-order axiomatization of the completely ordered field of real numbers ($\mathbf{R}^2$)—another classical result by Huntington [Hun02].[18] Thus, a natural response to the push-through argument is to make the move to second-order logic; after all, if our theory picks out just one isomorphism type, then the second application of the argument leading to radical indeterminacy is blocked.

We shall not provide an extensive treatment of the literature on categoricity, for it is rather vast. However, we do intend to give some glosses of how the discussion and research on categoricity touches on the quest for arithmetical determinacy.

1.5.1 Can second-order logic do the job?

As we have just mentioned, the following theorem could grant us **arithmetical determinacy**, at least up to the structural level:

Theorem 2 (Dedekind)**.** *All models of* $\mathbf{PA}^2$ *are isomorphic.*

Now, there is an important problem here, associated to the very use of second-order logic: it generates suspicion. To begin with, Quine famously said that second-order logic is just 'set theory in sheep's clothing' [Qui70]. What he was trying to get at is that second-order quantification is not really quantification over properties or relations, but quantification over sets; and this gets us too close to first-order set theory. But, if we need set theory to secure arithmetical determinacy, and we lack

[18] Exactly which complete second-order theories are categorical is still an open question. All in all, some results have been obtained. For instance, Solovay [Sol06] showed that any second-order, finitely axiomatizable theory is categorical assuming the axiom $V = L$; and [WV21] have shown that, assuming the axiom of projective determinacy, any finitely axiomatizable theory with a countable model is also categorical.

the sort of categoricity result for set theory that motivates the determinacy of the latter, have we made any progress? As a response, Vann McGee [McG97] takes on the challenge and proves a categoricity theorem for *second-order* **ZFC** with urelements. However, as he himself notes, whoever had serious doubts about the use of second-order logic will probably remain suspicious of his result, insofar as it still makes use of *second-order* logic for his set theory with urelements.[19] Note also that the concerns with second-order logic need not be reduced to the one expressed by Quine. The current debate seems to have moved past Quine's 'nominalist scruples' for properties or relations [FI19, 761]. Nowadays, the suspicion often arises out of the so-called 'overgeneration argument'. The precise lesson to be drawn from this argument is at the heart of an open philosophical debate (see e.g. [FI19, Pas14]); however, theorists agree that it has to do with passing for logic what isn't. In particular, for treating as validities (in the logical sense) certain mathematical statements that are clearly not logical truths—a result, some say, that is a consequence of the problematic degree of entanglement between mathematics and second-order logic.

Furthermore, the identification with set theory and the entanglement with mathematics are not the only aspects that generate concern for the skeptic of second-order logic. We are here dealing with *full* second-order logic, that is, the second-order semantics obtained when we allow the second-order quantification to occur over all subsets of the domain we are considering. This is to be differentiated from *Henkin* semantics, in which quantification may be restricted to a proper subset of all subsets of the domain—often, the collection of all definable subsets. The key point is that some crucial results of first-order logic do not carry over to full second-order logic, including the completeness theorem, the compactness theorem and the Löwenheim-Skolem theorem.

Now, if we were to swallow this bitter pill, and proceed with full second-order logic despite the absence of these fundamental results, we can ask: will it then be enough? Not really, for we would fall prey of a very Putnamian worry. As it happens, Dedekind's theorem only goes through when our semantics are the full semantics for second-order logic. That is, the result does not obtain with Henkin semantics. And this is no surprise, for the Löwenheim-Skolem theorem and Gödel's incompleteness theorems hold for second-order logic with Henkin semantics. So

[19] McGee then articulates a thorough defense of arithmetical determinacy that has been rather influential and, as such, must be mentioned in a footnote at least. His argument is based on the notion of open-endedness. As he sees it, the way in which we get to learn arithmetic is such that we know that the induction schema must come out as true for every predicate we encounter, and in whichever way we extend our language. Schemes like induction are essentially open-ended. As a result, open-endedness guarantees that we can rule out non-standard models, by introducing predicates that apply only to the standard numbers, and confirming that induction does not hold of them.

we must ensure that our semantics, that is, our quantifiers, are interpreted in the full rather than in the Henkin way. Accordingly, the Putnamian worry notices that there is no immediate way of ensuring that, and any viable option will need to invoke *more theory*, hence being subject to the just-more-theory manoeuvre. Putnam himself was clear on this [Put80, 481], and a very pristine exposition of what is going on is found in Button and Walsh:[20]

> In order for [Dedekind's] Theorem to do the job [we want] it to, [we] must have ruled out the Henkin semantics for second-order logic; indeed, ..., [we] must have shown that full models are *preferable* to Henkin models. But the distinction between full models and Henkin models essentially invokes abstract *mathematical* concepts. And, so the worry goes, the distinction between full and Henkin models is *just more theory*, and hence up for reinterpretation. [BW18, 159]

In sum, we would need a nice story about how we can make full semantics the intended one—that is, a story that is not subject again to Putnamian worries. Otherwise, the move to second-order logic is wholly unfruitful: Dedekind's theorem applies no more.

1.5.2 Categoricity without (full) second-order logic

We have just seen how the appeal to second-order logic in the seach for categoricity runs into new and old issues alike. We also mentioned that theories formulated in either first-order logic or second-order logic with Henkin semantics are not categorical, at least not if they have an infinite model. But what if this was not entirely correct? In this subsection, we explore categoricity results that do not require us to employ full second-order logic.

We begin by presenting Meadows' 'first-order' categoricity result for first-order **PA** [Mea13b, §3]. Certainly, **PA** has non-isomorphic models but a 'fake' second-order meta-theory can produce what resembles a categoricity result. The process is as follows. One works in a multi-sorted meta-theory, that is, one that allows more than one kind of object in the domain. Each kind of object is denoted by a different sort of variable. In our case, we have objects of one kind, denoted by x, y, z, and objects of a new, second kind, denoted by X, Y, Z. Moreover, the multi-sorted approach requires adding a predicate (in our case, the binary predicate $\in$) in such

[20] A non-exhaustive list of authors who have put forward similar remarks includes: Field (who located the indeterminacy of second-order notion here) [Fie94], Parsons [Par08, §48], and Meadows [Mea13b].

a way that we do not let two objects of the same kind be related through this new predicate, i.e., $x \in y$ and $X \in Y$ are not well-formed, but $x \in X$ is. Finally, the meta-theory for Meadows' categoricity result is obtained by adding to **PA**: (1) a comprehension axiom in the extended language, although restricted to formulae of the original language; and (2) an induction axiom quantifying universally over objects of the second kind (see [Mea13b, 528] for details). This theory is often known as $\mathbf{ACA}_0$, and Meadows claims it to be first-order, despite the second-order appearance.[21]

The whole idea behind a multi-sorted theory is that the objects represented by capital letters will stand for *classes*, i.e., representations of arithmetical structures. This will allow $\mathbf{ACA}_0$ to speak about models of **PA**. We now say that a model $\mathcal{M}$ is well-founded iff for every class X such that $X \cap |\mathcal{M}| = \emptyset$, there exists a $<_{\mathcal{M}}$-least element of $X \cap |\mathcal{M}|$. With $\mathbf{ACA}_0$ as our meta-theory, the categoricity result is finally stated thus:

Theorem 3. *Any two well-founded models of* **PA** *are isomorphic.*

For all the merit that this result deserves, I am unsure of the extent to which it is actually helpful to address either the push-through argument or the argument from Löwenheim-Skolem. For one thing, it does not entail that our theory of arithmetic singles out one isomorphism type—only that we can single it out ourselves by discarding non-well-founded models. But the notion of well-foundedness itself is a complex mathematical concept, once again subject to Putnamian worries. That is, either we assume a prior grasp of the notion, which seems implausible; or we need additional theoretical machinery that rules out non-standard interpretations of well-foundedness, falling prey to Putnam's arguments again. In any case, and on behalf of Meadows, we must acknowledge that he never submits that his result constitutes an appropriate response to Putnam.

A second way to obtain something that looks like categoricity without appealing to (full) second-order logic is what has been called *internal categoricity*. There are different understandings and ways of fleshing it out, and here we will follow the work done by Väänänen and Wang [VW15].[22] In this paper, the authors set to show that a form of categoricity for $\mathbf{PA}^2$ also holds for second-order logic with Henkin semantics. The way to proceed with second-order Henkin logic is to add to the

[21] For a standard exposition of $\mathbf{ACA}_0$, see e.g. [Sim09, ch.3].

[22] In a much more recent paper, Väänänen [V2̈1] proves an internal categoricity result for first-order **PA**. Considerations of space, and of technical complexity of the material, as well as the difficulty in interpreting what the theorem exactly shows, has prevented us from including it here. Nonetheless, it looks like an interesting line of research too.

logical axioms for second-order logic a comprehension axiom (CA) of the sort

$\exists X \forall \vec{x}(X(\vec{x}) \leftrightarrow \varphi(\vec{x}))$, for φ a second-order formula not containing X free.

Henkin logic thus defined is complete with respect to Henkin models, that we shall denote by pairs $(\mathcal{M}, \mathcal{G})$, although we will skip the details. For reference, see [VW15, 122].

The idea behind internal categoricity of this sort is to consider a language $\mathcal{L}$ that contains two copies of the language of arithmetic, i.e., $\mathcal{L} = (0, S, 0', S')$.[23] Let $PA^2(x, y)$ be an abbreviation of the conjunction in the language $\mathcal{L}' = (x, y)$ of the axioms of $\mathbf{PA}^2$. Then, internal categoricity amounts to the idea that a Henkin model formulated in such a language can 'see' the $\mathbf{PA}^2$-structures generated by each of the two copies as isomorphic. That is, if

$$(\mathcal{M}, \mathcal{G}) \vDash PA^2(S, 0) \wedge PA^2(S', 0')$$

then

$$(\mathcal{M}, \mathcal{G}) \vDash \exists R \, \mathrm{ISO} \, (R, (S, 0), (S', 0')).$$

In the formula above, ISO (R, x, y) formalizes the claim that R is an isomorphism between x and y.

The theorem is then the following:

Theorem 4 (Väänänen and Wang)**.**

$$CA \vdash (PA^2(S, 0) \wedge PA^2(S', 0')) \rightarrow \exists R \, ISO \, (R, (S, 0), (S', 0'))$$

How to understand this notion of internal categoricity exactly is not entirely clear yet. In any case, this line of research *prima facie* promising. After all, what led us to abandon second-order logic was the inability to decide which kind of semantics we 'live in'. But since *some* form of categoricity result obtains either way (that is, whether the semantics are full or Henkin), we may be entitled to appeal to second-order logic in order to establish **arithmetical determinacy**.

[23] Although Väänänen and Wang employ an additional predicate N (and the corresponding N') that stands for 'is a natural number'.

1.6 Alternative ways of securing determinacy

The last part of this chapter is dedicated to introducing two arguments in favour of **arithmetical determinacy** that have been less discussed in the literature than the model-theoretic arguments and the categoricity arguments. The first, due to Halbach and Horsten, can be seen as an alternative to arguments based on physical constraints *à la* Berry and Field in order to contest the argument from Löwenheim-Skolem.

1.6.1 Recursiveness and Tennenbaum's theorem

Volker Halbach and Leon Horsten [HH05] have advanced a position that they call (formalist) computational structuralism. This account is meant to give an answer to the question of how we can single out the isomorphism type corresponding to the standard model of arithmetic. Hence the 'structuralist' flavour, insofar as traditional structuralism defends that arithmetic is the study of *the* natural number structure. Drawing on a proposal in [Ben65], they defend that the way to rule out non-standard models of arithmetic has to do with the recursiveness of the relation $<$ and the operations of addition and multiplication. The motivation and justification for this line of reasoning comes from the following result, due to Tennenbaum [Ten59]:

Theorem 5 (Tennenbaum)**.** *Let $\mathcal{M}$ be a countable model of **PA**. If $+^{\mathcal{M}}$ is recursive, then $\mathcal{M}$ is isomorphic to the standard model of arithmetic.*

With Tennebaum's theorem in mind, they propose a thesis that describes how to identify a standard model. Let's write 'recursive*' for the practical, perhaps informal way of understanding recursiveness, i.e., an operation defined on certain symbols or notations is recursive* if there is a set of instructions that indicate how to manipulate those symbols on the basis of that operation. Then the thesis runs as follows [HH05, 183]:

> **REC**: Intended models are notation systems with recursive* operations on them satisfying the Peano axioms.

Now, let us clarify what this means. Initially, Halbach and Horsten propose that a model is intended or standard iff $<$, addition and multiplication are recursive in the theoretical sense (i.e., as belonging to a class of functions on $\mathbb{N}$). The problem with this account is that it assumes that the domain of the model will consist of the natural numbers. But what if the domain of the model we are assessing contains sets? Or sequences of strokes? As a consequence, they appeal to *coding*, the powerful machinery Gödel first came up with in order to allow formal systems to 'talk' about their own syntax. The idea is to do something similar with the domain of the ostensibly standard model: the model will be standard iff one can code the objects and appropriate relations of that model as, respectively, the natural numbers and the recursive* relations $<, +, \times$. At this point, however, they notice that what we really want coding to show us is that the model we are assessing picks up the appropriate structure. It is not a matter of there being a one-to-one (coding) relation between the domain in question and some strange, static objects called 'natural numbers', but between the domain and a set of symbols that displays the structure of the natural numbers 'in reality', together with the appropriate recursiveness*. Halbach and Horsten's final move consists then in going wholly formalist and structuralist: something counts as an intended model insofar as it presents some form of notation system (this can be Roman numerals, or von Neumann sets, or what have you) that satisfies *the* structure—the Peano structure—while having recursive* operations.

Formalist computational structuralism is meant to respond to the argument from Löwenheim-Skolem thus: should the advocate of the argument claim that we cannot tell whether we live in a non-standard model of arithmetic or not (for theoretical and observational constraints will not settle it down), one would just respond by pointing out to the fact that, after all, we can. We just need to check the recursiveness* of the operations and relations that are defined on the model.

All in all, the computational structuralist view has also received some criticism. Thus, Button and Smith [BS12] argue that it will not convince a Putnamian skeptic for a simple reason: a *practical* notion of recursiveness will not take us far enough. Since the computational structuralist needs to cover the totality of the operations of addition and multiplication, they would have to talk about what can be done in principle. But if they do it, then they are actually talking about what can be done in any arbitrary finite number of steps. So one needs to have a sharp conception of finite number beforehand, which is precisely what the argument from Löwenheim-Skolem disputes. To the best of my knowledge, it is still an open question whether the computational structuralist can adequately reply to this concern. And even if they did, recent results question the resort to Tennenbaum's theorem in order to secure the computational structuralist's thesis: Fedor Pakhomov [Pak22] has shown how to construct a theory that is definitionally equivalent to **PA** and which has a recursive model whose corresponding **PA**-model is not isomorphic to the standard model of arithmetic. Therefore, it may be that Tennenbaum's theorem does not apply with the generality that the computational structuralist needs.

1.6.2 Hyper-computers and supertasks

This kind of argument addresses the worries raised in section 1. To recap quickly, the idea there was that the existence of **PA**-independent sentences implies that these are indeterminate, and so that arithmetic is indeterminate. To this, the argument from hyper-computers replies that, should we have a hyper-computer that can check all contentious claims for every single natural number, we are justified in believing the truth or falsity of the statement, depending on the output of the hyper-computer.

Let's consider two examples: Goldbach's conjecture and the Gödel sentence for **PA**. Both are of the form $\forall n A(n)$, with $A(n)$ a Δ_0 sentence, that is, a sentence that contains no quantifiers other than bounded quantifiers. In sum, they are both Π^0_1 sentences. In the case of Goldbach's conjecture, we do not know it to be independent of **PA**, while the Gödel sentence is certainly independent. For Goldbach's conjecture, $A(n)$ will be a formalization of 'if n is greater than 2, then n is the sum of two primes'; for the Gödel sentence, $A(n)$ will be a more complicated number-theoretic statement. Now, certain solutions to the equations of general relativity seem to show that the existence of a device that can check an infinite number of steps in a finite amount of time (a 'supertask') is possible. This device, together with the conditions that allow it to perform the supertask, are sometimes known in the literature as a Malament-Hogarth machine. Then, the argument proceeds by letting the Malament-Hogarth machine check $A(n)$ on every $n \in \mathbb{N}$. For instance, if there exists some m for which $\neg A(m)$, the program could halt; and it will finalise running otherwise. At this point, one would offer some argument as to why we are justified in believing the outcome of the supertask, and why it counts as evidence that $\forall n A(n)$ (or $\neg\forall n A(n)$) is true. Sharon Berry [Ber14] has produced one such argument. As a conclusion, one would be entitled to claim that we have settled the determinate truth-value of Goldbach's conjecture, or the Gödel sentence, without appealing to any model-theoretic reasoning that might be subject to Putnamian worries. In fact, since *every* **PA**-independent statement of arithmetic is Π^0_1, then one could claim that arithmetic is in fact determinate, for each problematic statement will receive a determinate truth-value.

Due to limitations of space, we will not assess the argument from hyper-computers here. In any case, one shall note that many things must come together for the argument to succeed. First, one needs to argue convincingly for the possibility of Malament-Hogarth machines; the fact that there exist solutions to the equations of general relativity where these can be postulated does not guarantee that these can be actualised.[24] Secondly, one also needs a sound argument to the effect that we are justified in believing the outcome of the hyper-computer. And, finally, one needs a strong reasoning as to why believing the outcome of the hyper-computer is enough to attribute a determinate truth-value to the mathematical statement.[25] So, after all, making an argument from hyper-computers succeed in securing the determinacy of arithmetic is no easy task.

1.7 Final remarks

The reader will, by now, be familiar with the discussion around arithmetical determinacy. The lesson I would like them to extract is twofold: that our conviction to the determinacy of arithmetic urges us to meet the metasemantic challenge, and hence to explain how the determinacy arises; and that meeting this challenge is not straightforward, for we have arguments—most notably the model-theoretic arguments—that threaten determinacy in the first place. Moreover, the responses to Putnam's argument are either unsatisfactory (like Bueno's objection shows of Field's and Berry's arguments), or under-explored (like internal categoricity arguments). That is why we still need to live up to the task imposed by the metasemantic challenge.

[24] See [MR22, §2.3] for a summary of objections to the actual existence of Malament-Hogarth machines.

[25] Warren and Waxman [WW20b] have challenged the possibility of any such argument.

Chapter 2

As it should be clear by now, this project is a quest for mathematical determinacy; in particular, for arithmetical determinacy. But in this chapter, we take a step back to look at the broader picture. To be precise, we look at the logical picture. Very roughly, the idea is the following: in order to explore the question of mathematical determinacy from within the mathematical theories themselves, we need to implement a theory of *determinate truth*. This chapter is a first step towards designing one such theory. Since we first need to understand the virtues that we desire for our theory, and supervaluational-like truth (that is, truth that includes *all* logical truths) is one of them, we study different theories that display this kind of truth.

2.1 How to be supervaluational: three recipes

The notion of mathematical—and hence arithmetical—determinacy is intimately related to truth. As we explained in Chapter 1, determinacy for a subject matter is understood as there being a determinate truth-value for all the statements of the subject matter. It is a *semantic* notion of determinacy. It is no wonder then that we may want to look at the treatment the notion of truth receives within mathematical logic, in what is often known as *formal theories of truth*. These theories constitute a logical framework with which to understand what truth is, including (or particularly!) in the mathematical context. They are the means by which a mathematical theory can, in the very own object language in which the theory is formulated, speak about the truth of its statements. Hence, they stand as one of the essential elements of our investigation.

Actually, as we just mentioned, the ultimate goal of this chapter and others that will feature in this project is to produce a theory of truth. Or, rather, a theory of determinate truth. This will allow us to pose the very question on determinacy that we want to answer from within the mathematical theories on behalf of which we want to answer it. In doing so, however, we must be careful. The introduction of a truth predicate governed by a theory of truth often brings about as many problems as the ones the theory purported to solve. Semantic paradoxes are the nightmare as much as the engine of research in formal theories of truth. Therefore, in coming up with our theory, we aim for one that fares equally well in addressing the determinacy of purely mathematical statements and in accounting for the manners in which the notion of truth and that of determinate truth are handled. For instance, we want to be able to predicate truth of sentences that already contain a truth predicate; we want the notion to apply to sentences that include negations; and so on.[1]

The research on formal theories of truth is vast and very technical. Here, we would like to focus on theories that build on Saul Kripke's fixed-point semantics [Kri75]. It is also in the context of these theories that the formal notion of determinacy has been more prominently discussed. We cite just some examples. Solomon Feferman, in discussing his well-known system **KF**, developed an axiomatic theory of determinacy with axioms governing a rudimentary determinacy predicate [Fef91];[2] and, in a later paper [Fef08], he defended the axiomatic system **DT** for determinate truth. Hartry Field's 'Saving Truth From Paradox' [Fie08] is a pharaonic attempt to construct a theory of truth that escapes the revenge paradoxes associated with the concept of truth-theoretic determinacy, paradoxes that many Kripke-based theories of truth fall prey to. More recently, Volker Halbach and Kentaro Fujimoto [HFng] have advanced an axiomatic system for determinate truth in classical logic. Likewise, there have been attempts to work on notions similar to that of determinacy, a very popular one being that of 'groundedness'—see e.g. the work of Thomas Schindler [Sch14] and Lucas Rosenblatt [Ros21].

Now, to understand the motivation for this chapter, we note the following observation: one of the inconveniences that many fixed-point-based theories face lies in their inability to present as deteminate many truths of logic. Thus, we begin with the assumption that, say, 'If A, then A' should be considered a determinate truth for any sentence A under a reasonable theory of truth. But this is not the case under many Kripkean approaches. For instance, consider Kripke's own theory, and in particular the fixed point that many theorists consider most natural (even if this is not claimed by Kripke): the minimal fixed point with truth-value gaps. Then,

[1]The first condition rules out theories like Tarski's **CT** (for Compositional Truth) [Tar36]. Both the first and second condition rule out theories like **UTB** (for Uniform Tarski Biconditionals).

[2]When I say 'rudimentary' here, I just mean that the predicate is modelled after a very basic and intuitive understanding of what being determinate means. I have carried out some work where I argue that there are different and (arguably) more satisfying ways of formally understanding determinacy.

not all sentences of the form $A \rightarrow A$ fall in the extension of the truth predicate; and, consequently, no theory of determinacy formulated on top of this theory of truth will recover these sentences either—unless, of course, they are ready to claim that these sentences are determinately true but not true, which seems prima facie absurd. For this reason, we fix our eyes on supervaluationism.

Supervaluational theories of truth ensure that all sentences that are theorems of logic fall in the extension of the truth predicate. By 'logic', we just usually mean first-order logic. Supervaluationism is mostly associated with the application of supervaluatonist semantics to the Kripkean fixed-point construction. Thus, instead of the usual Strong (or Weak) Kleene logic, the satisfaction relation is formulated in supervaluationist terms. Details will follow in the next section. However, this is not the only way to get supervaluational. Recently, Johannes Stern [Ste18] has shown that one can obtain what he calls 'supervaluation-style truth' without being committed to van Fraassian semantics. Likewise, any theory of truth S whose truth predicate satisfies, meta-theoretically,[3] Tarski's T-Convention ($S \vdash \varphi$ iff $S \vdash \mathrm{T}\ulcorner\varphi\urcorner$), will also be supervaluational in spirit as long as the theorems of first-order logic are also theorems of S. A great example of that, which we will be studying for inspiration, is McGee's theory of truth [McG91]. The lesson is then that there are different ways to achieve supervaluational truth, and one does not necessarily need to stick to the van Fraassen-Kripke approach. Since we would like our prospect theory to be supervaluational too, this chapter will study the relations between these three paths to supervaluational truth.

On top of that, the fact that we take McGee's theory as part of our study has to do with the fact that it presents a very desirable feature for a theory of truth: material adequacy, i.e., the meta-theoretic satisfaction of Convention-T. We would like to be able to accommodate material adequacy as one of the features of our theory of truth and determinate truth, alongside supervaluationism.

Thus, this chapter will trace connections between supervaluatonist semantics *à la* Kripke-van Fraassen, Stern's proposal, and McGee's theory. Many of the results here are available in other papers, including Stern's and Cantini [Can90], but we also offer some new observations. Since we have a special interest in the virtues of McGee's theory, we dedicate a good deal of the chapter to its understanding; section 3 is entirely devoted to that. We must mention that, for this chapter, we have been unable to fully understand the depths of McGee's theory; as we will explain in due course, we have only worked out the details of one version of the theory. We hope

[3] As is well-known, any minimally strong theory from the logical point of view will be unable to satisfy Convention-T object-theoretically, due to Tarki's undefinitability result.

to accomplish this task in the near future.

Prior to section 3, we examine van Fraassian supervaluationist semantics (section 1) and Stern's supervaluational-style truth and its connections with Kripke-van Fraassen semantics and Cantini's theory **VF** (section 2). Section 4 traces further connections between Stern's theory and McGee's theory.

2.2 Supervaluationist theories of truth

Supervaluationist semantics was introduced by van Fraassen [vF66], in an attempt to deal with non-referring terms. Van Fraassen's goal—and, arguably, the goal of all supervaluationist semanticists—is to produce a semantics that respects what are now called *penumbral* truths, or truths that hold in virtue of (first-order) logic alone, even for sentences that do not receive a classical truth-value. Soon after [vF68], he considered an application of this semantics to the paradoxes of self-reference. Then, in his [Kri75], Kripke suggested that the fixed-point construction he was defining over a satisfaction relation based on Strong Kleene logic could be equally built using van Fraassen's supervaluations.

Let us quickly present the notational preliminaries before we set to explaining supervaluationist semantics. The conventions follow the line of [Hal14]. We will always be working with languages whose logical symbols are $\neg, \vee, \wedge, \forall, \exists$, as well as brackets. We also write $\varphi \rightarrow \psi$ as an abbreviation for $\neg\varphi \vee \psi$. As a default practice, and unless otherwise stated, we work with the language of arithmetic, $\mathcal{L}_0$ with its standard signature: $\{0, S, +, \times\}$. The language will be appropriately extended when needed, e.g., with a truth predicate. We will also assume that, since the theories we will be dealing with can interpret basic arithmetic, we can code the expressions of the language $\mathcal{L}_0$ and its extensions in some way—see e.g. [BBJ07]. $\bar{n}$ stands for the numeral of the number n (although we omit the bar for specific numbers). We write $\ulcorner\phi\urcorner$ for the numeral of the code of ϕ. As regards the truth-predicate, one would normally write the code of the sentence to which it applies in brackets, i.e., $\mathrm{T}(\ulcorner\phi\urcorner)$. However, to ease readability, we often drop the brackets when we use the upper corners. We shall also be able to code primitive recursive operations, and we represent them with a dot under them—e.g. $\underset{.}{\vee}, \underset{.}{\forall}, \underset{.}{\neg}$. There are a couple of exceptions, though. One is the substitution function, that we write as $x^{t}\!/\!_{s}$, for the result of substituting s with t in x. The other is the evaluation function: we write $\mathtt{val}(x)$ for the valuation of x. Given a language $\mathcal{L}$, we write $\mathrm{CT}_{\mathcal{L}}(x)$ to indicate that x is a closed term of $\mathcal{L}$, and $\mathrm{Sent}_{\mathcal{L}}(x)$ to indicate that x is a sentence of $\mathcal{L}$. We frequently omit the subscript $_{\mathcal{L}}$ if the language we work with is clear from the context. Finally, $\dot{x}$ will be the function that takes each number n to the numeral $\bar{n}$.

So much for the notation, and back to supervaluationism. Van Fraassen's initial idea was, in informal terms, the following: to capture the logical truths, a supervaluation over a model would consider the class of possible oscillations of the interpretation function for the truth-valueless atomic formulae, and assign the **T** truth-value to those sentences that come out true in every member of the class. When it comes to truth, the basic idea is identical: for an interpretation of the truth predicate, one considers the class of all interpretations that extend the given one. In his 1968 paper, van Fraassen opted for considering not the whole class of possible interpretation functions for the non-referring terms (respectively, the class of all interpretations of the truth predicate that extend the given one), but only those interpretations that are consistent in that they do not assign the same truth value to A and $\neg A$, for A a sentence of the language in question (respectively, interpretations of the truth predicate that extend the given one in a consistent way, i.e., such that $S \cup S^- = \varnothing$, for S^- the antiextension of the extended interpretation S). This is what he called an *admissible* valuation. In connection with the truth predicate, Kripke considered a further admissibility condition, namely, that the class of extended extensions of the truth predicate only includes extensions that are maximally consistent sets of sentences. Consider the language with the truth-predicate $\mathcal{L}_0^+ = \mathcal{L}_0 \cup \{\mathrm{T}\}$. Let capital letters S, R, or X, Y, Z, be sets of (codes of) sentences of $\mathcal{L}_0^+$, and let Greek letters such φ, ψ stand for formulae of $\mathcal{L}_0^+$. Let's write $X \vDash \varphi$ as short for $(\mathbb{N}, X) \vDash \varphi$, i.e., the classical model over the standard model of arithmetic where we let X be the extension of the truth predicate T.[4] Then, for we can thus distinguish three supervaluation schemes:[5]

$$S \vDash_{sv} \varphi \Leftrightarrow \forall Y \supseteq X (Y \vDash \varphi)$$
$$S \vDash_{vc} \varphi \Leftrightarrow \forall Y (Y \supseteq X \,\&\, Y \cap Y^- = \varnothing \Rightarrow Y \vDash \varphi)$$
$$S \vDash_{mc} \varphi \Leftrightarrow \forall Y (Y \supseteq X \,\&\, Y \in \mathrm{MAXCONS} \Rightarrow Y \vDash \varphi)$$

Here, if X is the extension of T, we write X^- for the antiextension (i.e., as in the above footnote, $\{\varphi | X \vDash \mathrm{T}\ulcorner\neg\varphi\urcorner\}$); and we write MAXCONS for the set of codes of all maximally consistent sets of sentences. There is a further supervaluation scheme,

[4]When we deal with partial evaluations, such $\vDash_{sv}$, $\vDash_{mc}$, and so on, it is normally expected that not only the extension but also the anti-extension of the truth predicate accompany the satisfaction relation. That is, it should be written $(\mathbb{N}, (X, X^-) \vdash \varphi$, for φ a sentence. But since we understand here the anti-extension as the set $\{\varphi | X \vDash \mathrm{T}\ulcorner\neg\varphi\urcorner\}$, we can omit reference to it every time we make use of the satisfaction relation.

[5]The labels we use here are borrowed from [FHKS15].

which Burgess [Bur86] attributes to van Fraassen—but that, as far I can tell, appeared for the first time in Burgess' paper only:

$$S \vDash_{vb} \varphi \Leftrightarrow \forall Y (Y \supseteq X \,\&\, Y \cap X^{-} = \varnothing \Rightarrow Y \vDash \varphi)$$

Hence, note that the diference between the vb and the vc supervaluation schemes lies in that the former requires the extended extension to be consistent with the initial antiextension, and the latter requires the extended extension to be consistent with its corresponding antiextension. Therefore, the class of extensions considered in vc is a subset of the class of extensions considered in vb, and so

$$S \vDash_{vb} \varphi \Rightarrow S \vDash_{vc} \varphi.$$

The Kripke jump operator based on the supervaluation schemes is then defined in the usual way as follows:

$$\mathcal{J}_e(S) := \{\ulcorner \varphi \urcorner \in \mathrm{Sent} |\, S \vDash_e \varphi\}, \text{ for } e \text{ one of: sv, vb, vc or mc.}$$

As usual, $\mathcal{J}_e(S) = S$ means that S is a fixed-point of $\mathcal{J}_e$. Moreover, the operators are monotone (i.e. $S \subseteq S' \Rightarrow \mathcal{J}_e(S) \subseteq \mathcal{J}_e(S')$), so the usual Woodruff-Martin-Kripke result guarantees that there will be fixed points for them. We often require that $S \in$ CONS, i.e., the set of consistent sets of sentences. In the case of the schemas vb, vc, and mc, this is particularly salient: otherwise, one application of the jump operator alone results in the degenerate fixed point (Sent, Sent). This is not necessarily so in the case of the sv operator; nonetheless, starting with an inconsistent set will certainly result in inconsistent fixed points even with this operator, and hence, in internally inconsistent fixed-point models.

The relations between the operators are as follows:

Proposition 1. *Let sv, vb, vc, mc and $\mathcal{J}_e(S)$ be as before. Then:*

i. For all sets $X \subseteq \omega$,

$$\mathcal{J}_{sv}(X) \subseteq \mathcal{J}_{vb}(X) \subseteq \mathcal{J}_{vc}(X) \subseteq \mathcal{J}_{mc}(X)$$

ii. For all sets X such that $X \in$ CONS and $X \notin$ MAXCONS,

$$\mathcal{J}_{sv}(X) \subsetneq \mathcal{J}_{vb}(X) \subsetneq \mathcal{J}_{vc}(X) \subsetneq \mathcal{J}_{mc}(X)$$

Proof. i follows from the way $S \vDash_e \varphi$ is defined for e each of the supervaluation schemes: just as with vb and vc, the class of extended extensions considered in the sv scheme is a superset of the class of extended extensions considered in the vb scheme, and the same holds for the vc scheme with respect to the mc scheme. For ii, one considers the following sentences, from left to right in the formula displayed: $\neg \mathrm{T}\ulcorner 0 \underset{\cdot}{=} 1 \urcorner$; $\forall x \neg(\mathrm{T}(x) \wedge \mathrm{T}(\underset{\cdot}{\neg} x))$; and $\forall x(\mathrm{T}(x) \vee \mathrm{T}(\underset{\cdot}{\neg} x))$. □

Supervaluationist theories are notoriously difficult to axiomatize. Thus, on the basis of considerations of complexity, Fischer et al. [FHKS15] proved the following result:

Theorem 6 (Fischer & Halbach & Kriener & Stern)**.** *Let e be an evaluation scheme such that $\mathcal{J}_{sv} \subseteq \mathcal{J}_e \subseteq \mathcal{J}_{mc}$. Then there is no recursively enumerable theory Σ such that*

$$(\mathbb{N}, S) \vDash \Sigma \Leftrightarrow \mathcal{J}_e(S) = S$$

This negative result proves the failure for Kripke's theory based on supervaluations of what Fischer et al. call '$\mathbb{N}$-categoricity'. An axiomatic theory T is $\mathbb{N}$-categorical with respect to an evaluation scheme f precisely when Theorem 6 fails for e, i.e., when $(\mathbb{N}, S) \vDash T \Leftrightarrow \mathcal{J}_f(S) = S$.[6]

It is often believed that $\mathbb{N}$-categoricity is key for an axiomatic theory to be an adequate axiomatization of a semantic theory of truth. If so, the supervaluational theories of truth we have surveyed cannot be axiomatized adequately. Yet Cantini [Can90] produced a theory that comes close to an axiomatization of the supervaluation scheme vc. The theory, known as **VF** and formulated in the language $\mathcal{L}_0^+$, consists of all the axioms of **PA** extended to all formulae of $\mathcal{L}_0^+$ (**PAT**), and the following axioms:

VF1 $\mathrm{T}\ulcorner A \urcorner \rightarrow A$

VF2 $\mathrm{CT}(x) \wedge \mathrm{CT}(y) \rightarrow (\mathrm{T}(x \underset{\cdot}{=} y) \leftrightarrow \mathtt{val}(x) = \mathtt{val}(y) \wedge \mathrm{T}(x \underset{\cdot}{\neq} y) \leftrightarrow \mathtt{val}(x) \neq \mathtt{val}(y))$

VF3 $\forall x(\mathrm{Ax}_{\mathbf{PAT}}(x) \rightarrow \mathrm{T}(x))$

VF4 $\forall z \mathrm{T}(x\, \dot{z}/v) \rightarrow \mathrm{T}(\underset{\cdot}{\forall} v x)$

VF5 $\mathrm{T}(x) \rightarrow \mathrm{T}(\ulcorner \mathrm{T}(\dot{x}) \urcorner)$

VF6 $\mathrm{Sent}(x) \wedge \mathrm{T}(\underset{\cdot}{\neg} \ulcorner \mathrm{T}(\dot{x}) \urcorner)) \rightarrow \mathrm{T}(\underset{\cdot}{\neg} x)$

VF7 $\mathrm{T}(x \underset{\cdot}{\rightarrow} y) \rightarrow (\mathrm{T}(x) \rightarrow \mathrm{T}(y))$

[6] **KF** is an example of one such theory, with e being the Strong Kleene evaluation scheme.

VF8 $\mathrm{T}(\dot{\neg}\ulcorner \mathrm{T}(\dot{x}) \wedge \mathrm{T}(\dot{\neg}\dot{x})\urcorner)$

VF9 $\mathrm{T}(\ulcorner \mathrm{T}(\dot{x}) \rightarrow \mathrm{Sent}(\dot{x})\urcorner)$

Here, $\mathrm{Ax}_{\mathbf{PAT}}(x)$ stands for 'x is an axiom of **PAT**'. While, once again, **VF** cannot $\mathbb{N}$-categorically axiomatize the semantic theory based on the scheme vc, Cantini proved that we get, at least, one of the directions behind $\mathbb{N}$-categoricity:

Theorem 7 (Cantini)**.** *Let X be such that $X \in$ CONS and $X = \mathcal{J}_{vc}(X)$. Then $X \vDash$* **VF**.

That is why we say that **VF** comes close to being an axiomatization of the supervaluational theory of scheme vc.

Moreover, Cantini showed that **VF** is, proof-theoretically, a very strong theory. For the lower bound, he showed

Theorem 8 (Cantini)**.** $\mathbf{ID}_1^{acc}$ *is interpretable in* **VF**.

And for the upper bound

Theorem 9 (Cantini)**.** **VF** *is interpretable in* **KPU**.

The arithmetical theorems of $\mathbf{ID}_1^{acc}$ and the arithmetical theorems **KPU** are known to be the same, namely, those of the theory $\mathbf{ID}_1$. When this happens, we say that the theories are proof-theoretically equivalent, which we write as $\mathbf{VF} \equiv \mathbf{ID}_1$. As mentioned, $\mathbf{ID}_1$ is a rather strong theory; in fact, it is one of the simplest examples of an impredicative theory (see [Poh09]). We note that $\mathbf{ID}_1$ is proof-theoretically equivalent to $(\Pi^1_1\text{-}\mathbf{CA})^-_0$ [Poh09, ch.13].[7]

One can equally formulate axiomatic theories $\mathbf{VF}^-$ and **VFM** such that Theorem 7 holds for them in relation to the supervaluation schemes vb and mc. In particular, $\mathbf{VF}^-$ contains axioms VF1-VF7 and VF9 from above (we can call them VF^-1-VF^-8), and **VFM** contains axioms VF2-VF9 (call them VFM1-VFM8) and the axiom

VFM9 $\mathrm{T}\ulcorner \mathrm{Sent}(\dot{x}) \rightarrow \mathrm{T}(\dot{\neg}\dot{x}) \vee \mathrm{T}(\dot{x})\urcorner$

Lemma 1. *Let X be such that $X \in$ CONS.*

i) If $X = \mathcal{J}_{vb}(X)$, then $X \vDash$ $\mathbf{VF}^-$.

[7] $(\Pi^1_1\text{-}\mathbf{CA})^-_0$ is the theory that includes the basic axioms of all subsystems of second-order arithmetic for $+, \times, <$; the induction axiom $(0 \in X \wedge \forall n(n \in X \rightarrow n+1 \in X) \rightarrow \forall n(n \in X))$; and the comprehension scheme $(\exists X \forall n(n \in X \leftrightarrow \varphi(n)))$ restricted to formulae φ that are Π^1_1 and that contain no free set parameters.

ii) If $X = \mathcal{J}_{mc}(X)$, then $X \vDash$ **VFM**

Proof. We just need to follow Cantini's proof for Theorem 7. In the case of i): first note that, since $X \subseteq X$, $X \in$ CONS implies that $X \vDash_{vb} A \Rightarrow X \vDash A$.

- VF$^-$1: Assume $X \vDash \mathrm{T}\ulcorner A \urcorner$. Then $A \in X$, and so $X \vDash_{vb} A$, so $X \vDash A$.
- VF$^-$2: Assume $X \vDash \mathrm{CT}(x) \wedge \mathrm{CT}(y)$. Then

  $\mathbb{N} \vDash \mathtt{val}(x) = \mathtt{val}(y)$
  $\Leftrightarrow X \vDash_{vb} x = y$ (since this would be satisfied in all models of $\mathbb{N}$)
  $\Leftrightarrow \ulcorner x = y \urcorner \in \mathcal{J}_{vb}(X) = X$
  $\Leftrightarrow \mathcal{J}_{vb}(X) = X \vDash \mathrm{T}(x \underset{\cdot}{=} y)$
- VF$^-$3: Let $X \vDash \mathrm{Ax}_{\mathbf{PAT}}(x)$. Then, if $x = \ulcorner A \urcorner$, then $Y \vDash A$ for all $Y \supseteq X$ (for A holds in every $\mathbb{N}$-model over $\mathcal{L}_T$). So $X \vDash_{vb} A$, hence $\mathcal{J}_{vb}(X) = X \vDash \mathrm{T}\ulcorner A \urcorner$.
- VF$^-$4: Assume $X \vDash \forall z \mathrm{T}\ulcorner x^{\dot{z}}/_{v} \urcorner$. Let $x = \ulcorner A \urcorner$. For X' such that $\mathcal{J}_{vb}(X') = X$, we have $X' \vDash_{vb} A^{\bar{n}}/_{v}$ for each $n \in \omega$, and so $X' \vDash_{vb} \forall v A$. Hence, $X \vDash_{vb} \forall v A$, so $\mathcal{J}_{vb}(X) = X \vDash \mathrm{T}\ulcorner \forall v A \urcorner$.
- VF$^-$5: Assume $X \vDash \mathrm{T}(x)$:

  $X \vDash \mathrm{T}(x)$
  $\Leftrightarrow x \in Y$ for all $Y \supseteq X$
  $\Leftrightarrow X \vDash_{vb} \mathrm{T}(x)$
  $\Leftrightarrow \mathcal{J}_{vb}(X) = X \vDash \mathrm{T}\ulcorner \mathrm{T}(\dot{x}) \urcorner$
- VF$^-$6: Assume $X \vDash \neg \mathrm{T}(\underset{\cdot}{\neg} x)$. Let $x = \ulcorner A \urcorner$. Then, $\ulcorner \neg A \urcorner \notin X$. So there are models $(\mathbb{N}, Y)$ with $Y \supseteq X$ and $Y \cap X^- = \varnothing$ such that $Y \vDash \mathrm{T}\ulcorner A \urcorner$. Therefore, $X \nvDash_{vb} \neg \mathrm{T}\ulcorner A \urcorner$, so $\ulcorner \neg \mathrm{T}\ulcorner A \urcorner \urcorner \notin \mathcal{J}_{vb}(X) = X$, hence $X \vDash \neg \mathrm{T}\ulcorner \neg \mathrm{T}\ulcorner A \urcorner \urcorner$.
- VF$^-$7: Just as with VF$^-$5, $X \vDash \mathrm{T}(x \underset{\cdot}{\rightarrow} y)$ and $X \vDash \mathrm{T}(x)$ imply $X \vDash_{vb} A \rightarrow B$ and $X \vDash_{vb} A$, so $X \vDash_{vb} B$—where $x = \ulcorner A \urcorner$ and $x = \ulcorner B \urcorner$.
- VF$^-$8: with the definition $\mathcal{J}_{vb}$, we can prove by transfinite induction that $X \subseteq$ Sent for any X such that $\exists X'(\mathcal{J}_{vb}(X') = X)$. Then, X being a fixed point, $X \vDash \mathrm{T}(x)$ implies $x \in$ Sent. Therefore, $\mathbb{N} \vDash \mathrm{Sent}(x)$, and so $X \vDash \mathrm{Sent}(x)$.

For ii). We can use the same arguments as above for VFM1-VFM7. The equivalent of VF8, VFM8, follows from the fact that the supervaluation relation mc is defined in terms of $Y \subseteq X$ such that $Y \in$ MAXCONS $\subseteq$ CONS (indeed, $\neg(\mathrm{T}(x) \wedge \mathrm{T}(\underset{\cdot}{\neg} x))$ for any x gets in the extension of the truth predicate after just one application of the operator, whatever the initial set X'). The truly different axiom is VFM9, but this follows from the supervaluation scheme mc too. If $X \vDash \mathrm{Sent}(x)$, then $\mathbb{N} \vDash \mathrm{Sent}(x)$, and also $Z \vDash \mathrm{Sent}(x)$ for all Z; and, just like before, $\mathrm{T}(x) \vee \mathrm{T}(\underset{\cdot}{\neg} x)$ for any x gets in the extension of the truth predicate after just one application of the operator, whatever the initial set X'. □

It is no surprise that VF1 is not an axiom of **VFM**. Otherwise, **VFM** would be inconsistent—by well-known results of Friedmand and Sheard [FS87], all theories satisfying axioms VF1 and VFM9 are. However, VF1 cannot be proved under the mc scheme because the latter is not 'classically sound', in the sense that $X \vDash_{mc} A$ does not imply $X \vDash A$—not even if X is a fixed point of the scheme.

Let's finally add some proof-theoretic considerations for $\mathbf{VF}^-$. Just like **VF**, the latter is proof-theoretically equivalent to $\mathbf{ID}_1$. Following Theorem 9, and since $\mathbf{VF}^-$ is a subtheory of **VF**,

Proposition 2. $\mathbf{VF}^-$ *is interpretable in* **KPU**.

The other direction follows from a result by Friedman and Sheard [FS87], also employed by Stern:

Proposition 3. $\mathbf{VF}^-$ *proves all arithmetical theorems of Bar Induction* (**BI**).

Proof. By a theorem in [FS87], we know that this result holds for any theory of truth whose truth predicate satisfies the principles below. As it is clear, $\mathbf{VF}^-$ is one such theory.

1. $\mathrm{T}\ulcorner A \urcorner \rightarrow A$

2. $\forall z \mathrm{T}(x\,{}^{\dot{z}}/_{v}) \rightarrow \mathrm{T}(\underset{\cdot}{\forall} v x)$

3. $\mathrm{T}(x \underset{\cdot}{\rightarrow} y) \rightarrow (\mathrm{T}(x) \rightarrow \mathrm{T}(y))$

4. $\mathrm{T}\ulcorner A \urcorner$ for A an axiom of **PAT**.

□

BI is known to have the same arithmetical theorems as $\mathbf{ID}_1$, hence $\mathbf{VF}^- \equiv \mathbf{VF} \equiv \mathbf{ID}_1$—recall the notation above. The proof-theoretic strength of **VFM** remains an open problem.

2.3 Stern's supervaluation-style truth and IT

Johannes Stern [Ste18] has recently argued that the supervaluation schemes present a major drawback with two undesirable consequences. The drawback in question is that, instead of a fully compositional approach to truth, the scheme relies on a rather intransparent process which renders equally intransparent constructions. The first, most evident consequence is that the resulting fixed-point theories are not compositional (e.g. they do not validate the principle $\mathrm{T}(x \dot{\vee} y) \leftrightarrow \mathrm{T}(x) \vee \mathrm{T}(y)$). The second is that the intransparency extinguishes any chance of providing a neat proof-theoretic axiomatization. As we mentioned, there is no $\mathbb{N}$-categorical axiomatization that will do the job.

Accordingly, Stern has proposed an evaluation scheme (actually, two) and its accompanying axiomatic theory (theories) that (partially) remedy these concerns. We say partially because the resulting theories are not compositional in the strict sense of the word. For instance, they do not validate the compositional principle for disjunction either. But they do allow us (or so does Stern claim) to locate the failure of compositionality. Moreover, the evaluation scheme is more tractable (as demonstrated by the lower complexity of the fixed points generated).

We sketch Stern's idea briefly. The supervaluational evaluation $\vDash_e$, for $e \in \{\mathrm{sv}, \mathrm{vb}, \mathrm{vc}, \mathrm{mc}\}$ for a given extension of the truth predicate S seems to work on two levels: on the one hand, it gathers the set of sentences S' that the Strong Kleene (SK) scheme satisfies in the S-model; on the other, it accounts for penumbral truths by gathering all sentences that follow from S' in the intersection of a class of models that meet a given condition. The problem is that, in doing so, the schemes go well beyond first-order logic (and beyond second-order consequence as well—see [Ste18, 824]).[8] But the idea behind supervaluationism was to respect *first-order* logical truths (see [vF66, 484]), so Stern proposes an evaluation schema that is based on, and does not go over, the first-order consequence relation. The scheme also operates on two levels, for a given extension S of the truth predicate: it collects the set S' SK-satisfied by the model where S is the extension of T; and then adds the set of sentences that are first-order satisfied in every model that makes true S' and **PAT** (so that the set of penumbral truths also includes the theorems of number theory). After an appropriate simplification, the schema, labelled as SSK (for Supervaluation Strong Kleene) can be defined as follows. We still operate in the language $\mathcal{L}_0^+$,

[8] A result that buttresses this claim has been obtained by Kremer and Urquhart [KU08]: they show that the set of existential validities of second-order logic (Π^1_2 − SOL) is recursively encodable in the satisfaction relation that obtains from the supervaluation scheme sv.

of which φ, ψ are formulae and S is a set of (codes of) sentences:

$$S \vDash_{ssk} \varphi \Leftrightarrow \exists\psi(S \vDash_{sk} \psi \,\&\, \mathbf{PAT} \vdash \psi \rightarrow \varphi)$$

Then, we can define the (monotone) jump operator just like before:[9]

$$\mathcal{J}_{ssk}(S) := \{\varphi \,|\, S \vDash_{ssk} \varphi\}$$

Because the relation involved is notably less complex than on the supervaluational case, Stern shows that we can define an arithmetical operator $\Theta(x)$ such that

Theorem 10 (Stern). $F_\Theta = F_{ssk}$

F_Θ (respectively, F_{ssk}) is the set of fixed points of the Θ operator (respectively, $\mathcal{J}_{ssk}$), i.e., the set of all X such that $\Theta(X) = X$ (respectively, $\mathcal{J}_{ssk}(X) = X$). This arithmetical operator closely resembles the one that yields fixed-points for T under the SK scheme. Actually, we need to define the arithmetic formula $\xi(x, X)$ that forms the basis of that operator first, in order to introduce Θ. Note that we omit writing $\mathrm{T}^{\mathbb{N}}$ everytime we write T.

$$\begin{aligned}
\xi(x,X) := \quad & x \in \mathrm{True}_0 && (2.1)\\
\vee \quad & \exists y, z(x = (y \dot{\vee} z) \wedge (y \in X \vee z \in X)) && (2.2)\\
\vee \quad & \exists y, z(x = (y \dot{\wedge} z) \wedge (y \in X \wedge z \in X)) && (2.3)\\
\vee \quad & \exists y, v(x = (\dot{\forall} v y) \wedge \forall z(y^{z}/_{v} \in X)) && (2.4)\\
\vee \quad & \exists y, v(x = (\dot{\exists} v y) \wedge \exists z(y^{z}/_{v} \in X)) && (2.5)\\
\vee \quad & \exists t(x = (\dot{\mathrm{T}}(t)) \wedge \mathtt{val}(t) \in X) && (2.6)\\
\vee \quad & \exists t(x = (\dot{\neg}\dot{\mathrm{T}}(t)) \wedge (\dot{\neg}\mathtt{val}(t)) \in X \vee \neg\mathrm{Sent}(\mathtt{val}(t))) && (2.7)
\end{aligned}$$

As we said, this is the formula that yields an operator whose fixed points are the Strong-Kleene fixed points. On the basis of $\xi(x, X)$, we now define Stern's operator Θ:

$$\begin{aligned}
\Theta(X) := \{n \in \omega \,|\, \mathbb{N} \vDash \quad & \xi(n, X) && (2.8)\\
& \vee \exists x(\xi(x, X) \wedge \mathrm{Pr}_{\mathbf{PAT}}(x \dot{\rightarrow} n))\} && (2.9)
\end{aligned}$$

[9]For consistency, we stick to the notation we employed before and do not follow Stern's here.

Moreover, Stern introduces an analogue of the supervaluatonist scheme vc. As we saw, the latter is characterised by the fact that it only considers models where the extension and the anti-extension of T are disjoint. The main consequence of this is that, for any x,

$$\ulcorner\neg(\mathrm{T}(\dot{x}) \wedge \mathrm{T}(\underset{\cdot}{\neg}\dot{x}))\urcorner \in \mathcal{J}_{vc}(X)$$

for all sets X. So Stern just 'forces' this in the definition of this second scheme, that he labels SSK_c. Write $\mathrm{con}(\phi)$ if there exists t, s such that $t^{\mathbb{N}} = s^{\mathbb{N}}$ and ϕ is of the form $\neg(\mathrm{T}(t) \wedge \mathrm{T}(\underset{\cdot}{\neg}s))$. Then

$$S \vDash_{ssk_c} \varphi \Leftrightarrow \exists\psi((S \vDash_{sk} \psi \vee \mathrm{con}(\psi))\,\&\,\mathbf{PAT} \vdash \psi \rightarrow \varphi)$$

An arithmetical operator Θ_c can be easily provided by replacing the disjunct (2.8) in the formula above with $\xi_c(x, X) := \xi(x, X) \vee \mathrm{con}(x)$, and replacing $\xi(x, X)$ with $\xi_c(x, X)$ in disjunct (2.9). Theorem 10 will also hold for SSK_c and Θ_c. Furthermore, Stern shows that the minimal SSK-fixed point (I_{ssk}) coincides with the minimal vb-fixed point (I_{vb}), as do I_{ssk_c} and I_{vc}; but the schemes are not equivalent, for not all fixed points coincide. More interestingly perhaps, the semantic theories **SSK** and $\mathbf{SSK}_c$ admit of an $\mathbb{N}$-categorical axiomatisation. These are the theories **IT** and $\mathbf{IT}_c$. **IT** comprises the axioms of **PAT** plus the following:

IT1 $\mathrm{CT}(x) \wedge \mathrm{CT}(y) \rightarrow (\mathrm{T}(x \underset{\cdot}{=} y) \leftrightarrow \mathtt{val}(x) = \mathtt{val}(y))$

IT2 $\mathrm{CT}(x) \wedge \mathrm{CT}(y) \rightarrow (\mathrm{T}(x \underset{\cdot}{\neq} y) \leftrightarrow \mathtt{val}(x) \neq \mathtt{val}(y))$

IT3 $\forall x, y(\mathrm{Sent}(x \underset{\cdot}{\wedge} y) \rightarrow (\mathrm{T}(x) \wedge \mathrm{T}(y) \rightarrow \mathrm{T}(x \underset{\cdot}{\wedge} y)))$

IT4 $\forall x, y(\mathrm{Sent}(x \underset{\cdot}{\vee} y) \rightarrow (\mathrm{T}(x) \vee \mathrm{T}(y) \vee \exists z(\xi(z, \mathrm{T}) \wedge \mathrm{Pr}_{\mathbf{PAT}}(z \underset{\cdot}{\rightarrow} x \underset{\cdot}{\vee} y) \leftrightarrow \mathrm{T}(x \underset{\cdot}{\vee} y)))$

IT5 $\forall v, x(\mathrm{Sent}(\underset{\cdot}{\forall} v x) \rightarrow (\forall z \mathrm{T}(x\, \dot{z}/v) \rightarrow \mathrm{T}(\underset{\cdot}{\forall} v x)))$

IT6 $\forall x, y(\mathrm{Sent}(\underset{\cdot}{\exists} v x) \rightarrow (\exists z \mathrm{T}(x\, \dot{z}/v) \vee \exists w(\xi(w, \mathrm{T}) \wedge \mathrm{Pr}_{\mathbf{PAT}}(w \underset{\cdot}{\rightarrow} \underset{\cdot}{\exists} y)) \leftrightarrow \mathrm{T}(\underset{\cdot}{\exists} v x)))$

IT7 $\forall x(\mathrm{T}(x) \rightarrow \mathrm{T}\ulcorner\mathrm{T}(\dot{x})\urcorner)$

IT8 $\forall t(\mathrm{T}(\underset{\cdot}{\neg}\mathtt{val}(t)) \vee \neg\mathrm{Sent}(\mathtt{val}(t)) \leftrightarrow \mathrm{T}\ulcorner\neg\mathrm{T}(t)\urcorner)$

IT9 $\forall x, y(\mathrm{T}(x) \wedge \mathrm{Pr}_{\mathbf{PAT}}(x \underset{\cdot}{\rightarrow} y) \rightarrow \mathrm{T}(y))$

IT10 $\forall x(\mathrm{T}(\underset{\cdot}{\neg} x) \rightarrow \neg\mathrm{T}(x))$

IT11 $\mathrm{T}\ulcorner\forall x(\mathrm{T}(\dot{x}) \rightarrow \mathrm{Sent}(\dot{x}))\urcorner$

IT12 $\forall t_1, ..., t_n(\mathrm{T}\ulcorner\varphi(t_1, ..., t_n)\urcorner \to \varphi(\mathtt{val}(t_1), ..., \mathtt{val}(t_n))$

For $\mathbf{IT}_c$, we substitute IT4 and IT6 for

IT4* $\forall x, y(\mathrm{Sent}(x \dot{\vee} y) \to (\mathrm{T}(x) \vee \mathrm{T}(y) \vee \exists z(\xi(z, \mathrm{T}) \vee \mathrm{con}(z) \wedge \mathrm{Pr}_{\mathbf{PAT}}(z \dot{\to} x \dot{\vee} y) \leftrightarrow \mathrm{T}(x \dot{\vee} y)))$

IT6* $\forall x, y(\mathrm{Sent}(\dot{\exists} \upsilon x) \to (\exists z \mathrm{T}(x(\dot{z}/\upsilon)) \vee \exists z(\xi(z, \mathrm{T}) \vee \mathrm{con}(z) \wedge \mathrm{Pr}_{\mathbf{PAT}}(z \dot{\to} \dot{\exists} y)) \leftrightarrow \mathrm{T}(\dot{\exists} \upsilon x)))$

and we also add

ITC $\forall x(\mathrm{T}\ulcorner\mathrm{T}(\dot{\neg}\dot{x}) \to \neg\mathrm{T}(\dot{x})\urcorner)$

As announced, **IT** and $\mathbf{IT}_c$ are $\mathbb{N}$-categorical with respect to the semantic theories defined by the schemes SSK and SSK_c, that is:

Theorem 11 (Stern). *For any $S \in CONS$,*

$$S \models \mathbf{IT} \Leftrightarrow \mathcal{J}_{ssk}(S) = S$$
$$S \models \mathbf{IT}_c \Leftrightarrow \mathcal{J}_{ssk_c}(S) = S$$

Just like there are deep connections between Stern's SSK/ SSK_c and the van Frassian supervaluational schemes, **IT** and Cantini's **VF** present interesting points in common as well. The first thing to notice is that **VF** is a subtheory of $\mathbf{IT}_c$.[10] On top of that, the three theories (**IT**, $\mathbf{IT}_c$, and **VF**) are proof-theoretically equivalent in the sense of proving the same arithmetical theorems (see above):

Theorem 12 (Stern). $\mathbf{VF} \equiv \mathbf{IT} \equiv \mathbf{IT}_c \equiv \mathbf{ID}_1 \equiv \Pi^1_1\text{-}\mathbf{CA}_0^- \equiv \mathbf{KPU}$

2.3.1 Stern's theory and dependent truth

This subsection draws attention to a further, interesting fact that connects Stern's theory with a theory of truth developed by Hannes Leitgeb and revised by Toby Meadows. Thus, in his [Lei05], Leitgeb argues that the kind of sentences that we can plausibly apply the T-schema to are those who ultimately 'depend on' another sentence from the language without the truth predicate. These are, precisely, the sentences traditionally conceived as grounded. Accordingly, he constructs a theory of truth based on the introduction of a dependence operator. Some time later,

[10] VF7 is not listed among the axioms of **IT**, but it can be seen to follow from IT3 and IT9.

Meadows [Mea13a] refined Leitgeb's theory in order to account for the groundedness of sentences like $\mathrm{T}\ulcorner 0 = 0\urcorner \vee \lambda$ or $\mathrm{T}\ulcorner 0 = 1\urcorner \wedge \lambda$ (where λ is the Liar). The version we briefly sketch is thus Meadows'. Let a set $Y \subseteq$ Sent be a *safe expansion* of a set X (notation: $Y \sqsupseteq X$) iff (i) $Y \supseteq X$, and (ii) $X \cap \{\ulcorner\neg\varphi\urcorner | \ulcorner\varphi\urcorner \in Y\} = \varnothing$. Then, for φ a formula of $\mathcal{L}_0^+$, and X, Y, Z sets of codes of sentences of the language:

Definition 1 (Meadows)**.** *φ is double-conditional dependent on X given Z (notation: $\varphi \dashv_Z X$) iff for all $Y \sqsupseteq Z$, $(\mathbb{N}, Y) \vDash \varphi \Leftrightarrow (\mathbb{N}, X \cap Y) \vDash \varphi$.*

Now, we define two intertwined fixed-point processes:

Definition 2 (Leitgeb, Meadows)**.**

Hierarchy of readiness:	*Jump operator:*
$\Phi_0 = \varnothing$	$\Gamma_0 = \varnothing$
$\Phi_{\alpha+1} = \{\varphi \mid \varphi \dashv_{\Gamma_\alpha} \Phi_\alpha\}$	$\Gamma_{\alpha+1} = \{\varphi \in \Phi_{\alpha+1} \mid (\mathbb{N}, \Gamma_\alpha) \vDash \varphi\}$
$\Phi_\xi = \bigcup_{\alpha<\xi} \Phi_\alpha$*, for ξ a limit ordinal*	$\Gamma_\xi = \bigcup_{\alpha<\xi} \Gamma_\alpha$*, for ξ a limit ordinal*

The idea behind the two hierarchies goes back to the notion of dependence: a formula is 'ready' to be evaluated by the truth operator when it essentially depends on the extension of the truth predicate at the previous stage of the operator. All grounded sentences will be ready eventually, so all will get evaluated at some point in the sequence of Γ_α . Moreover, since the operator is monotonic, this sequence also generates a fixed-point in the sense of Kripke. Given that it will be the minimal fixed point (for we started with $\Gamma_0 = \varnothing$), we call this fixed point I_Γ. Meadows then showed the following result:

Theorem 13 (Meadows)**.** *$I_\Gamma = I_{vb}$.*

Recall that I_{vb} is the minimal fixed point obtained over the supervaluation scheme vb. Meadows [Mea13a, 239] finds the identity of the minimal fixed points rather appealing. As he sees it, the supervaluation scheme offers a simple, mathematically elegant and well-understood way to obtain a theory of truth (perhaps contra Stern's claim), whereas Leitgeb's theory of dependent truth is better motivated philosophically. Therefore, the fact that the minimal fixed points of both theories are the same means that this set of truths unites the best of both approaches. Be that as it may, and since, as we mentioned, Stern showed that $I_{ssk} = I_{vb}$, it is evident that we also get the following identity:

Corollary 1. *$I_\Gamma = I_{vb} = I_{ssk}$*

While more research can be done here, it is interesting to notice that the three minimal fixed points coincide. It is not known to us whether all fixed points of Γ and the scheme vb coincide, but we certainly know that this is not the case with vb and **SSK**. For all we know, it could also be that the fixed points of Γ and **SSK** coincide. In any case, and without that information to hand, it might well be that this equivalence is telling us something about the privileged position of the minimal fixed point when it comes to supervaluation-like fixed points.

2.4 McGee's theory of definite truth

In this section, we sketch McGee's theory of definite truth as presented in [McG91, ch. 8], and *describe* the theory, in the sense of writing down the principles obeyed by its truth predicate.

Let us begin with a disclaimer. McGee's theory is, in fact, two theories. One is presented in [McG91, Th. 8.8], the other one in [McG91, Th. 8.13].[11] This chapter will only deal with the second one; unfortunately, as of today, I am still in the process of understanding the complexities behind the first theory. The truth is that the additions McGee introduces for the second theory make it much more tractable than the first. We are committed to a full grasp of that theory for the near future of the project.

Since, as mentioned, there are two theories at stake, and we want to focus on the second one, we will call it **MG2**. But before we expound the theory, let's mention some preliminaries and necessary terminology.

A partial interpretation is a pair $(\mathcal{M}, \Gamma)$: $\mathcal{M}$ is some ordinary first-order model in some language $\mathcal{L}$, and Γ is a set of sentences in a language $\mathcal{L}^+ \supseteq \mathcal{L}$. The partiality lies in the fact that, while $\mathcal{L}$ is fully interpreted, $\mathcal{L}^+$ is not—there will be predicates whose extension is not fully determined. The motivation behind this is to account for vague predicates and, most relevantly, for the truth predicate—a predicate McGee considers vague too. Now, McGee's task is to show how one can provide a recursive set of sentences over a partial interpretation such that (i) the resulting theory is a conservative extension of the partial interpretation and (ii) the theory is a materially adequate theory of truth for the initial partial interpretation together with the set of sentences of the theory.[12] In the case we are interested in,

[11] There is a third theory, but this adds nothing to the meat of the truth predicate; it just helps to work out a predicate for *definite* truth.

[12] In other words: the theory of truth must be materially adequate for sentences with the truth predicate too.

namely a theory of truth over the standard model of arithmetic $\mathbb{N}$ with no partially interpreted predicates other than the truth predicate, he shows that we can obtain a theory $(\mathbb{N}, X)$ such that $(\mathbb{N}, X)$ is a conservative extension of $\mathbf{PA}_\omega$ and such that X is a materially adequate theory of truth for $(\mathbb{N}, X)$. As we mentioned rather informally above, the latter means that we can find a formula $\tau(x)$ such that, for all formulae of $\mathcal{L}_0^+$ (as defined in the previous sections):

$$(\mathbb{N}, X) \models_{\mathrm{D}} \varphi \text{ iff } (\mathbb{N}, X) \models_{\mathrm{D}} \tau(\ulcorner \varphi \urcorner)$$
$$\text{and}$$
$$(\mathbb{N}, X) \models_{\mathrm{D}} \neg\varphi \text{ iff } (\mathbb{N}, X) \models_{\mathrm{D}} \neg\tau(\ulcorner \varphi \urcorner)$$

When this is the case, we say that $\tau(x)$ is a materially adequate truth predicate (MATP). Here, $\models_{\mathrm{D}}$ is the relation of *proof-theoretic definite truth*, which is defined as follows:

Definition 3. *Given the standard model of arithmetic* $\mathbb{N}$*, and a partial interpretation* $(\mathbb{N}, \Gamma)$*, we write* $(\mathbb{N}, \Gamma) \models_{\mathrm{D}} \varphi$ *(read:* φ *is definitely true in* $(\mathbb{N}, \Gamma)$ *in the proof-theoretic sense) iff there is a derivation sequence that ends with* φ *and whose only members are:*

i) a member of Γ*.*

ii) an atomic or negated atomic sentence true in $\mathbb{N}$*.*

iii) an axiom of first-order logic.

iv) the result of applying modus ponens to previous members of the sequence.

v) the result of applying the ω*-rule to previous members of the sequence.*

Because in this section the only notion of provability we will be dealing with is the one just introduced, we will simply write $(\mathbb{N}, \Gamma) \vdash \varphi$ when we want to write $(\mathbb{N}, \Gamma) \models_{\mathrm{D}} \varphi$. We will recover this more elaborate notation later on.

If this much is clear, we can now introduce **MG2**. Let $\Delta \ulcorner \xi \urcorner$ be the following set of (codes of) sentences:

- All axioms of **KF**, formulated with the predicate 'Kr'.
- The formalisation of 'T is a maximal consistent set of sentences of $\mathcal{L}_1^+$', where $\mathcal{L}_1^+$ is the language obtained by adding the predicate T to the language $\mathcal{L}_1 := \mathcal{L}_0 \cup \{\mathrm{Kr}\}$.
- The formalisation of the sentence '$\forall$ sentences $x \in \mathcal{L}_1^+$ $[[\mathrm{Kr}(\ulcorner \xi(x) \urcorner) \wedge \forall$ finite, inconsistent set $R, \exists r \in R(\mathrm{Kr}(\ulcorner L(\xi(x), \xi(r)) \urcorner))] \to \mathrm{T}(x)]$'

Details of why we include these sentences as our theory of truth can be found in [McG91, ch.8]. In any case, here's a hint. The axioms of **KF** are essential to guarantee that **MG2** includes the minimal SK-fixed point, which acts as some form of universal set in that it 'contains' all inductive sets—and allows to define a MATP on the basis of that containment. One of the 'prices' we pay is, however, that our theory is formulated in a rather unnatural language, one that includes two truth predicates: Kr and T.

The two other sentences are the way to force the theory to be based only on models such that the extension of T is in MAXCONS. Regarding the first sentence, McGee offers no formalisation of it (and leaves it as 'the formalisation of...'). It seems to us that the most intuitive such formalisation will be something like:

$$\forall x(\mathrm{Sent}_{\mathcal{L}_1^+}(x) \to (\neg \mathrm{T}(x) \leftrightarrow \mathrm{T}(\dot{\neg} x))).$$

As per the last sentence, it ensures that, under the extension of T, we get as many sentences in a consistent way as possible; and it does so by selecting the largest consistent set of sentences that can be obtained by mirroring the construction of the inductive set **MG2** in the minimal fixed point. The idea is not to obtain a theory that includes a maximally consistent set of sentences;[13] rather, what lies behind is some sort of conviction that maximal consistency is a desirable feature of the models that make up the theory of truth, as long as they all share a core of truths.[14]

Now, and just for the record, I will briefly digress on the difficulties I have found to understand McGee's theories, in particular the theory that we have not presented here, and that we shall call **MG1**. **MG1** is just like **MG2** but, instead of $\Delta\ulcorner\xi\urcorner$, we just have the axioms of **KF**, formulated with the predicate Kr. Without going into detail, McGee exploits a trick from recursive set theory to come up with a formula that acts as a MATP (call it $\tau(x)$) for the closure of the axioms of **KF** under ω-logic. This formula τ, however, is defined in terms of various other formulae that capture provability, refutability, and some form of well-ordering of formulae, in the theory. The main problem we have faced is that these other formulae present a rather intricate behaviour across models, and dealing with them has turned out to be a difficult challenge. While these mechanisms at stake in **MG1** are still playing a basic role in the theory **MG2**, the addition of the two other sentences that complete $\Delta\ulcorner\xi\urcorner$ allows us to reason on the basis of the maximal consistency of T over the

[13] And in fact, it clearly doesn't. It isn't true that **MG2** $\vdash \mathrm{T}\ulcorner\varphi\urcorner$ or **MG2** $\vdash \mathrm{T}\ulcorner\neg\varphi\urcorner$ for all $\varphi \in$ Sent. As with the supervaluation scheme mc, **MG2** proves the disjunction in the object theory (**MG2** $\vdash \mathrm{T}\ulcorner\varphi\urcorner \vee \mathrm{T}\ulcorner\neg\varphi\urcorner$), not in the meta-theory.

[14] Thus, McGee [McG92] has defended that maximal consistency cannot be the sole or main driving principle in the construction of a theory of truth.

models of the theory, which hugely simplifies the task. Hence, the principles of what at first sight may appear like a more complex theory of truth (**MG2**), happen to be easier to extract than the principles of the simpler **MG1**.

Having finished the digression, let us now describe what the theory **MG2** looks like. Thus, we begin by proving the principles obeyed by the truth predicate of **MG2**. Theorem 8.13 of McGee shows that T is a materially adequate predicate for $(\mathbb{N}, \Delta\ulcorner\xi\urcorner)$. We now show that it also obeys the following principles:

Theorem 14. *For any closed terms $s, t \in \mathcal{L}_0$, and for any sentences $A, B \in \mathcal{L}_1^+$, $(\mathbb{N}, \Delta\ulcorner\xi\urcorner)$ proves the following principles:*

1. $\mathrm{T}(s \underset{\cdot}{=} t) \leftrightarrow \mathtt{val}(s) = \mathtt{val}(t))$
2. $\mathrm{T}(s \underset{\cdot}{\neq} t) \leftrightarrow \mathtt{val}(s) \neq \mathtt{val}(t)$
3. $\neg\mathrm{T}(s \underset{\cdot}{=} t) \leftrightarrow \mathtt{val}(s) \neq \mathtt{val}(t)$
4. $\mathrm{T}\ulcorner\neg A\urcorner \leftrightarrow \neg\mathrm{T}\ulcorner A\urcorner$ *(COMP)*
5. $\mathrm{T}\ulcorner\neg\neg A\urcorner \leftrightarrow \mathrm{T}\ulcorner A\urcorner$ *(DNEG)*
6. $\mathrm{T}\ulcorner A \wedge B\urcorner \leftrightarrow (\mathrm{T}\ulcorner A\urcorner \wedge \mathrm{T}\ulcorner B\urcorner)$ *(CONJ)*
7. $\mathrm{T}\ulcorner A \vee B\urcorner \leftrightarrow (\mathrm{T}\ulcorner A\urcorner \vee \mathrm{T}\ulcorner B\urcorner)$ *(DISJ)*
8. $\mathrm{T}\ulcorner\forall v A\urcorner \rightarrow \forall z \mathrm{T}\ulcorner A\dot{z}/v\urcorner$ *(UNIV$_2$)*
9. $\exists z \mathrm{T}\ulcorner A\dot{z}/v\urcorner \rightarrow \mathrm{T}\ulcorner\exists v A\urcorner$ *(EXIST$_2$)*
10. $\neg(\mathrm{T}\ulcorner A\urcorner \wedge \mathrm{T}\ulcorner\neg A\urcorner)$ *(CONS)*

While basically all proofs can be achieved by reasoning on the basis of the maximal consistency of the extension of T, abbreviated MAXCONS, we will proceed slightly differently. MAXCONS will only be needed for the following lemma:[15]

Lemma 2. $(\mathbb{N}, \Delta\ulcorner\xi\urcorner)$ *proves that, for all sentences $A, B \in \mathcal{L}_T$,*

$$(\mathrm{T}\ulcorner A\urcorner \wedge \mathrm{T}\ulcorner A \rightarrow B\urcorner) \rightarrow \mathrm{T}\ulcorner B\urcorner$$

Proof. We fist prove, in the standard fashion, that the claim holds when the extension of T is a maximally consistent set. We show that, given a MAXCONS set Δ, if φ is derivable in propositional logic from Δ (that we write as $\Delta \vdash_{PL} \varphi$), then $\varphi \in \Delta$. Assume not. Let $\Gamma := \Delta \cup \{\varphi\}$. Pick some arbitrary ψ such that $\Gamma \vdash_{PL} \psi$. Call $\mathcal{D}_1$

[15] McGee [McG91, 206] already points out that this result holds, although he offers no proof.

the derivation from Δ to φ, and the derivation from Γ to ψ $\mathcal{D}_2$. Then, $\mathcal{D}_1 \cup \mathcal{D}_2$ is a derivation from Δ to ψ. Being ψ arbitrary, we note that $\Gamma \vdash_{PL} \bot$ implies $\Delta \vdash_{PL} \bot$. Since the latter cannot be because Δ is MAXCONS, $\Gamma \nvdash_{PL} \bot$. But then, Δ is consistent with φ. So, by MAXCONS, $\varphi \in \Delta$. Finally, given the set $\Delta' := \{A, A \to B\}$, we have $\Delta' \vdash_{PL} B$.

Finally, we make use of the fact (established by theorem 7.1 in McGee) that, if φ holds in all models $(\mathbb{N}, X)$ such that $(\mathbb{N}, X) \vDash \Delta\ulcorner\xi\urcorner$, then $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash \varphi$. Indeed, the reasoning above shows that, given any such model $(\mathbb{N}, S)$, $(\mathbb{N}, S) \vDash \mathrm{T}\ulcorner A\urcorner \wedge \mathrm{T}\ulcorner A \to B\urcorner$ implies $(\mathbb{N}, S) \vDash \mathrm{T}\ulcorner B\urcorner$—hence, all models of $\Delta\ulcorner\xi\urcorner$ satisfy $(\mathrm{T}\ulcorner A\urcorner \wedge \mathrm{T}\ulcorner A \to B\urcorner) \to \mathrm{T}\ulcorner B\urcorner$. □

We are then set to prove to theorem 14:

Proof of Theorem 14.

For principles 1-3: We know that either $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash s = t$ or $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash s \neq t$ for all closed terms s, t of the language of arithmetic without T and Kr (i.e., $\mathcal{L}_0$). Moreover, $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash s = t$ iff $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash \mathrm{T}(s \dot{=} t)$, and $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash s \neq t$ iff $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash \neg\mathrm{T}(s \dot{=} t)$. Thus, for instance, the left to right direction of (1) just requires us to pick any arbitrary s, t and show $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash \neg\mathrm{T}(s \dot{=} t) \vee \mathtt{val}(s) = \mathtt{val}(t)$. We know that either $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash \neg\mathrm{T}(s \dot{=} t)$, in which case the disjunction follows, or $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \nvdash \neg\mathrm{T}(s \dot{=} t)$, from which $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \nvdash s \neq t$ and so $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash s = t$ follows. The other direction, as well as principles 2 and 3, are basically the same.

For 4: this follows from the fact that $\Delta\ulcorner\xi\urcorner$ includes the formalisation of 'T is a maximal consistent set of sentences of $\mathcal{L}_1^+$'. As we said, while McGee is not very clear on how this formalization looks like, it will be something equivalent to: $\forall x(\mathrm{Sent}_{\mathcal{L}_1^+}(x) \to (\mathrm{T}(\dot{\neg} x) \leftrightarrow \neg\mathrm{T}(x)))$. Therefore, if $\ulcorner A\urcorner \in \mathrm{Sent}_{\mathcal{L}_1^+}$, the claim obtains.

For the remaining cases, we make use of the fact that

$$(*) \quad (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash \mathrm{T}\ulcorner\varphi\urcorner \text{ for } \varphi \text{ an axiom of first-order logic.}$$

We reason within $(\mathbb{N}, \Delta\ulcorner\xi\urcorner)$ in every proof:

For 5: We prove one direction, the other one is symmetric.

1. $\mathrm{T}\ulcorner A \to \neg\neg A\urcorner$ (*)
2. $(\mathrm{T}\ulcorner A \to \neg\neg A\urcorner \wedge \mathrm{T}\ulcorner A\urcorner) \to \mathrm{T}\ulcorner\neg\neg A\urcorner$ Lemma 2
3. $\mathrm{T}\ulcorner A\urcorner \to \mathrm{T}\ulcorner\neg\neg A\urcorner$ 1,2; prop logic

For 6: We prove the right-to-left direction; the left-to-right is fairly straightforward with the axioms of logic $A \wedge B \to A$ and $A \wedge B \to B$.

1. $T\ulcorner A \to (B \to A \wedge B)\urcorner$ (*)
2. $T\ulcorner A\urcorner \wedge T\ulcorner A \to (B \to A \wedge B)\urcorner \to T\ulcorner B \to A \wedge B\urcorner$ Lemma 2
3. $T\ulcorner B\urcorner \wedge T\ulcorner B \to A \wedge B\urcorner \to T\ulcorner A \wedge B\urcorner$ Lemma 2
4. $T\ulcorner A \to (B \to A \wedge B)\urcorner \wedge T\ulcorner A\urcorner \wedge T\ulcorner B\urcorner \to T\ulcorner A \wedge B\urcorner$ 2,3; prop logic
5. $T\ulcorner A\urcorner \wedge T\ulcorner B\urcorner \to T\ulcorner A \wedge B\urcorner$ 1,4; prop logic

For 7: The right-to-left direction is almost immediate. The left-to-right is as follows:

1. $T\ulcorner A \vee B \to \neg(\neg A \wedge \neg B)\urcorner$ (*)
2. $T\ulcorner \neg(\neg A \wedge \neg B)\urcorner \to \neg T\ulcorner \neg A \wedge \neg B\urcorner$ Principle 4
3. $\neg T\ulcorner \neg A \wedge \neg B\urcorner \to \neg(T\ulcorner \neg A\urcorner \wedge T\ulcorner \neg B\urcorner)$ Principle 6; prop logic
4. $\neg(T\ulcorner \neg A\urcorner \wedge T\ulcorner \neg B\urcorner) \to \neg T\ulcorner \neg A\urcorner \vee \neg T\ulcorner \neg B\urcorner$ Prop logic
5. $\neg T\ulcorner \neg A\urcorner \vee \neg T\ulcorner \neg B\urcorner \to T\ulcorner A\urcorner \vee T\ulcorner B\urcorner$ Principles 4 and 5, prop logic
6. $T\ulcorner \neg(\neg A \wedge \neg B)\urcorner \to T\ulcorner A\urcorner \vee T\ulcorner B\urcorner$ 2,5; prop logic
7. $T\ulcorner A \vee B\urcorner \to T\ulcorner A\urcorner \vee T\ulcorner B\urcorner$ 1, 6; Lemma 2, prop logic

For 8: Note that there is only one direction here.

1. $T\ulcorner \forall v A \to A(\bar{n})\urcorner$ For each $n \in \omega$, (*)
2. $T\ulcorner \forall v A \to A(\bar{n})\urcorner \wedge T\ulcorner \forall v A\urcorner \to T\ulcorner A(\bar{n}))$ For each $n \in \omega$, Lemma 2
3. $T\ulcorner \forall v A\urcorner \to T\ulcorner A(\bar{n})\urcorner$ For each $n \in \omega$, 1,2; prop logic
4. $\forall x(T\ulcorner \forall v A\urcorner \to T\ulcorner A(\dot{x})\urcorner)$ 3, ω-rule
5. $T\ulcorner \forall v A\urcorner \to \forall x T\ulcorner A(\dot{x})\urcorner$ 4, FOL

Principle 9 is very straightforward too, with just an application of the axiom of FOL

$$A(\bar{n}) \to \exists v A.$$

Principle 10 follows from principle 4 immediately. $\square$

This, together with material adequacy, completes the principles of the truth predicate for **MG2**. We shall continue by noting which principles are inconsistent with it. But first, an observation. It seems to us that one lesson that can be extracted from Theorem 14 and Lemma 2 is the following: that any model of a theory based on principles 1-3, maximal consistency of T, and the principle (*) alone, will already obtain principles 5-10. This is, after all, what is stake in McGee's theory. The material adequacy of T within the theory, together with the agreement across models on what counts as a logical truth and as a true atomic or negated atomic sentence is what yields principles 1-3 and (*) for each model of the theory, whereas $\Delta\ulcorner\xi\urcorner$ (and, in particular, the second bullet point above) is what forces the extension of T to be maximally consistent at each model. But any other route that results in the same principles being satisfied by the models will do, since the remaining of the proof does not depend on any specific feature of **MG2**.

Now, we indicate which principles are inconsistent with the truth predicate of **MG2**. But first, we recall a well-known theorem by McGee [McG85]. The formulation of the theorem differs slightly from McGee's own, and is taken from [Ste17]:

Theorem 15 (McGee)**.** *Let Γ be a theory extending Robinson's arithmetic Q in a language $\mathcal{L}_T$ that includes some truth predicate* T*, that is closed under the rule*

$$\frac{\phi}{\mathrm{T}\ulcorner\phi\urcorner}\ \textit{T-Intro}$$

and proves

- $\mathrm{T}\ulcorner\neg\phi\urcorner \to \neg\mathrm{T}\ulcorner\phi\urcorner$ *(CONS)*
- $\forall x\mathrm{T}\ulcorner\phi(x)\urcorner \to \mathrm{T}\ulcorner\forall v\phi(v)\urcorner$ *(UNIV$_1$)*
- $\mathrm{T}\ulcorner\phi \to \psi\urcorner \to (\mathrm{T}\ulcorner\phi\urcorner \to \mathrm{T}\ulcorner\psi\urcorner)$ *(IMP)*

for all $\phi, \psi \in \mathcal{L}_T$. Then, Γ is ω-inconsistent, i.e. there is a formula χ such that $\Gamma \vdash \chi(\bar{n})$ for each $n \in \omega$ but $\Gamma \vdash \exists x\neg\chi(x)$.

Finally, the principles that are inconsistent with **MG2** are as follows:

Remark 1. $(\mathbb{N}, \Delta\ulcorner\xi\urcorner)$ is inconsistent with any of the following principles:

1. $\mathrm{T}\ulcorner\neg\mathrm{T}\ulcorner A\urcorner\urcorner \leftrightarrow \mathrm{T}\ulcorner\neg A\urcorner$
2. $\mathrm{T}\ulcorner\mathrm{T}\ulcorner A\urcorner\urcorner \leftrightarrow \mathrm{T}\ulcorner A\urcorner$

3. $\mathrm{T}\ulcorner A \urcorner \to A$

4. $A \to \mathrm{T}\ulcorner A \urcorner$

5. $\forall z \mathrm{T}\ulcorner A \dot{z}/v \urcorner \to \mathrm{T}\ulcorner \forall v A \urcorner$

6. $\mathrm{T}\ulcorner \exists v A \urcorner \to \exists z \mathrm{T}\ulcorner A \dot{z}/v \urcorner$

Proof. Principle 1: it can be divided into the two directions of the biconditional, that we shall treat as two principles. Both of them are inconsistent with CONS and material adequacy:

1. $\mathrm{T}\ulcorner \neg A \urcorner \to \mathrm{T}\ulcorner \neg \mathrm{T}\ulcorner A \urcorner\urcorner$, which we shall call ¬REP. Note that, here as in the following proofs, it will suffice to prove λ, $\neg\lambda$, $\mathrm{T}\ulcorner \lambda \urcorner$ or $\neg\mathrm{T}\ulcorner \lambda \urcorner$ and a contradiction can be reached easily with material adequacy. Here, λ is the Liar for T. We reason within $(\mathbb{N}, \Delta\ulcorner \xi \urcorner)$:

i.	$\mathrm{T}\ulcorner \neg\lambda \urcorner \to \mathrm{T}\ulcorner \neg \mathrm{T}\ulcorner \lambda \urcorner\urcorner$	¬REP
ii.	$\mathrm{T}\ulcorner \neg \mathrm{T}\ulcorner \lambda \urcorner\urcorner \to \mathrm{T}\ulcorner \lambda \urcorner$	def of λ
iii.	$\mathrm{T}\ulcorner \neg\lambda \urcorner \to \mathrm{T}\ulcorner \neg\lambda \urcorner$	
iv.	$\mathrm{T}\ulcorner \neg\lambda \urcorner \to \mathrm{T}\ulcorner \lambda \urcorner \wedge \mathrm{T}\ulcorner \neg\lambda \urcorner$	
vi.	$\neg(\mathrm{T}\ulcorner \lambda) \wedge \mathrm{T}\ulcorner \neg\lambda \urcorner)$	CONS
vii.	$\neg\mathrm{T}\ulcorner \neg\lambda \urcorner$	Modus tollens
viii.	$\neg\neg\lambda$	Material adequacy
ix.	λ	

2. We shall call this principle ¬DEL. The proof of its inconsistency runs as follows:

i.	$\mathrm{T}\ulcorner \lambda \urcorner \to \mathrm{T}\ulcorner \neg \mathrm{T}\ulcorner \lambda \urcorner\urcorner$	def of λ
ii.	$\mathrm{T}\ulcorner \neg \mathrm{T}\ulcorner \lambda \urcorner\urcorner \to \mathrm{T}\ulcorner \neg\lambda \urcorner$	by ¬DEL
iii.	$\mathrm{T}\ulcorner \lambda \urcorner \to \mathrm{T}\ulcorner \lambda \urcorner \wedge \mathrm{T}\ulcorner \neg\lambda \urcorner$	
iv.	$\neg(\mathrm{T}\ulcorner \lambda \urcorner \wedge \mathrm{T}\ulcorner \neg\lambda \urcorner)$	CONS
v.	$\neg\mathrm{T}\ulcorner \lambda \urcorner$	Modus tollens

Principle 2: it can be divided into two directions:

1. $\mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner$, which we shall call REP, and is inconsistent with CONS and material adequacy.

i. $T\ulcorner\lambda\urcorner \to T\ulcorner\neg\ulcorner T\ulcorner\lambda\urcorner\urcorner$ def of λ

ii. $T\ulcorner\lambda\urcorner \to T\ulcorner T\ulcorner\lambda\urcorner\urcorner$ by REP

iii. $T\ulcorner\lambda\urcorner \to T\ulcorner T\ulcorner\lambda\urcorner \wedge T\ulcorner\neg T\ulcorner\lambda\urcorner\urcorner$

iv. $\neg(T\ulcorner T\ulcorner\lambda\urcorner\urcorner \wedge T\ulcorner\neg T\ulcorner\lambda\urcorner\urcorner)$ by CONS

v. $\neg T\ulcorner\lambda\urcorner$ Modus tollens

2. $T\ulcorner T\ulcorner A\urcorner) \to T\ulcorner A\urcorner$, which we shall call DEL, and is inconsistent with CONS, DNEG and material adequacy.

i. $T\ulcorner T\ulcorner\lambda\urcorner\urcorner \to T\ulcorner\neg\neg T\ulcorner\lambda\urcorner\urcorner$ DNEG

ii. $T\ulcorner\neg\neg T\ulcorner\lambda\urcorner\urcorner \to T\ulcorner\neg\lambda\urcorner$ def of λ

iii. $T\ulcorner T\ulcorner\lambda\urcorner\urcorner \to T\ulcorner\neg\lambda\urcorner$

iv. $T\ulcorner T\ulcorner\lambda\urcorner\urcorner \to T\ulcorner\lambda\urcorner$ DEL

v. $T\ulcorner T\ulcorner\lambda\urcorner\urcorner \to T\ulcorner\lambda\urcorner\urcorner \wedge T\ulcorner\neg\lambda\urcorner$

vi. $\neg(T\ulcorner\lambda\urcorner \wedge T\ulcorner\neg\lambda\urcorner)$ CONS

vii. $\neg T\ulcorner T\ulcorner\lambda\urcorner\urcorner$ Modus tollens

viii. $\neg T\ulcorner\lambda\urcorner$ Material adequacy

Principle 3, sometimes known as T-OUT, is inconsistent with material adequacy by Montague's paradox (see e.g. [McG91, ch.1]).

Principle 4, known as T-IN, is also inconsistent with material adequacy—in fact, once again, it is inconsistent even with one direction of material adequacy, namely T-ELIM or CONEC. We can reason as follows; as before, we only prove up to $\neg\lambda$.

i. $\lambda \to T\ulcorner\lambda\urcorner$ T-IN

ii. $\lambda \to \neg T\ulcorner\lambda\urcorner$ def of λ

iii. $\neg\lambda$

Principle 5: its inconsistency is a direct consequence of theorem 15, by lemma 2 and principle 9 of theorem 14; if **MG2** is ω-inconsistent, it is inconsistent simpliciter, for it counts the ω-rule among its rules of inference.

Principle 6: We note how to derive principle 5 from principle 6.

1. $\neg\exists z T\ulcorner\neg A\dot{z}/v\urcorner \to \neg T\ulcorner\exists v\neg A\urcorner$ Principle 6, prop logic

2. $\forall z\neg T\ulcorner\neg A\dot{z}/v\urcorner \to T\ulcorner\forall z\neg\neg A\dot{z}/v\urcorner$ 1; prin. 4 in Theorem 14, prop logic, (*)

3. $\forall z \mathrm{T}\ulcorner \neg\neg A\dot{z}/v \urcorner \rightarrow \mathrm{T}\ulcorner \forall z \neg\neg A \dot{z}/v \urcorner$ 2; prin. 4 in Theorem 14, prop logic

4. $\forall z \mathrm{T}\ulcorner A\dot{z}/v \urcorner \rightarrow \mathrm{T}\ulcorner \forall z A \dot{z}/v \urcorner$ 3; prin. 5 in Theorem 14, prop logic

□

2.5 Relations between Stern's theories and MG2

In this section, we establish some relations between Stern's theory $\mathbf{SSK}_{mc}$ (an extension of **SSK** that can be seen as the counterpart of the supervaluatonist scheme mc—we shall explain more about it later) and McGee's theory **MG2**. In doing so, we also characterise McGee's theory further and provide some insights into how we can think about this theory in terms of inductive definitions. In particular, in section 4.1 we first show that **MG2** is the minimal fixed point of two different inductive definitions, and we go on to show that all fixed points of these inductive definitions coincide. We then use this feature to establish, in section 4.2, that the minimal fixed point of Stern's $\mathbf{SSK}_{mc}$ is a subset of McGee's **MG2**.

2.5.1 MG2 as the minimal fixed point of two inductive definitions

In theorem [McG91, Th. 7.3], McGee shows that, given any partial interpretation $(\mathbb{N}, \Gamma)$, the following inductive definition characterises the theory $\{\varphi | (\mathbb{N}, \Gamma) \models_{\overline{\mathrm{D}}} \varphi\}$ in the sense that $\{\varphi | (\mathbb{N}, \Gamma) \models_{\overline{\mathrm{D}}} \varphi\}$ is its minimal fixed point:[16]

$$
\begin{aligned}
\forall x[R(x) \leftrightarrow \{x \in Sent_{\mathcal{L}_1^+} \wedge \\
& [G(x) \\
& \vee x \text{ is an axiom of logic} \\
& \vee x \in \ \mathrm{True}_0 \\
& \vee \exists y, v(x = (\underset{\cdot}{\forall} v y) \wedge \forall z R(y\, \dot{z}/v)) \\
& \vee \exists y(R(y) \wedge R(y \underset{\cdot}{\rightarrow} x))]\}]
\end{aligned}
$$

Here, True_0 is the set of atomic or negated atomic formulae true in $\mathbb{N}$, and $G(x)$ is a formula whose extension is the set of (codes of) sentences in Γ. As a result, if we substitute $G(x)$ in this inductive definition for $D(x)$, a formula whose extension is

[16]Note that we have adopted the more cumbersome notation $(\mathcal{M}, \Gamma) \models_{\overline{\mathrm{D}}} \varphi$ again in this section.

the set of sentences $\Delta\ulcorner\xi\urcorner$, the resulting inductive definition characterises the theory **MG2** in the sense we mentioned, i.e., **MG2** is its minimal fixed point.

We prove the same result for a different inductive definition, based on the arithmetical operators that are often used to characterise axiomatic theories of truth, and in particular the one in [Ste18]. This inductive definition is the formula $\phi_{mc}(x, X)$:

$$
\begin{aligned}
\phi_{mc}(x,X) \Leftrightarrow\, & x \in \mathrm{Sent}_{\mathcal{L}_1^+} \wedge \\
& [[(D(x) \\
& \vee \exists y, z(x = (y \dot{\vee} z) \wedge (y \in X \vee z \in X) \\
& \vee \exists y, z(x = (y \dot{\wedge} z) \wedge (y \in X \wedge z \in X) \\
& \vee \exists y, v(x = (\dot{\forall} v y) \wedge \forall z(y \dot{z}/v \in X) \\
& \vee \exists y, v(x = (\dot{\exists} v y) \wedge \exists z(y \dot{z}/v \in X) \\
& \vee \exists t(x = (\dot{\mathrm{T}} t) \wedge \mathtt{val}(t) \in X) \\
& \vee \exists t(x = (\dot{\neg} \dot{\mathrm{T}} t) \wedge (\dot{\neg}\mathtt{val}(t)) \in X)] \\
& \vee \exists y(y \in X \wedge \mathrm{Pr}_{\mathbf{PAKT}}(y \dot{\rightarrow} x)) \\
& \vee \exists s, t(x = \ulcorner \mathrm{T}(s) \vee \mathrm{T}(\dot{\neg} t)\urcorner \wedge \mathtt{val}(t) = \mathtt{val}(s)) \\
& \vee \exists s, t(x = \ulcorner \neg(\mathrm{T}(s) \wedge \mathrm{T}(\dot{\neg} t))\urcorner \wedge \mathtt{val}(t) = \mathtt{val}(s))]
\end{aligned}
$$

In the formula above, D is again the predicate whose extension is the set of codes of the formulae in $\Delta\ulcorner\xi\urcorner$. $\mathrm{Pr}_{\mathbf{PAKT}}$ is a predicate expressing provability in **PAKT**, i.e. Peano Arithmetic over the language $\mathcal{L}_1^+$.

We can then formulate the following operator:

$$\Phi_{mc}(X) := \{x | \mathbb{N} \vDash \phi_{mc}(x, X)\}$$

As a matter of fact, we could have come up with a much shorter inductive definition that would also do the job. In particular, the following inductive definition presents the same fixed points as the one we have just introduced:

$$
\begin{aligned}
\phi_{mc^*}(x,X) \Leftrightarrow & x \in \mathrm{Sent}_{\mathcal{L}_1^+} \wedge \\
& [[(D(x) \\
& \vee \exists y, v(x = (\dot{\forall} v y) \wedge \forall z(y^{\dot{z}}/v \in X) \\
& \vee \exists t(x = (\dot{\mathrm{T}} t) \wedge \mathtt{val}(t) \in X) \\
& \vee \exists t(x = (\dot{\neg}\dot{\mathrm{T}} t) \wedge (\dot{\neg}\mathtt{val}(t)) \in X)] \\
& \vee \exists y(y \in X \wedge \mathrm{Pr}_{\mathbf{PAKT}}(y \dot{\rightarrow} x))]
\end{aligned}
$$

The two inductive definitions are not strictly the same—if we use this new definition to formulate an operator Φ_{mc^*} in the same way we set Φ_{mc}, we can see that $\Phi_{mc^*}(X) \neq \Phi_{mc}(X)$ for some X. In other words, the stages in the construction of the fixed points differ, *even if the fixed points themselves do not*. The key here is that, for the case of Φ_{mc^*}, it 'takes more stages' for many formulae with connectives to be included in the set that will constitute the fixed point than it would take under the operator Φ_{mc}. As an example, let's examine the case of conjunction. Suppose that $x, y \in X$, that X is not a fixed point of Φ_{mc} nor Φ_{mc^*}, and that $(x \dot{\wedge} y) \notin X$. The clause for conjunction in $\phi_{mc}(x,X)$ yields $(x \dot{\wedge} y) \in \Phi_{mc}(X)$ immediately. However, when it comes to the operator Φ_{mc^*}, the corresponding formula $\phi_{mc^*}(x,X)$ lacks a clause for conjunction, and so needs to proceed through the clause that captures provability in **PAKT**. Since when $x, y \in \mathrm{Sent}_{\mathcal{L}_1^+}$, we have $\mathbb{N} \models \mathrm{Pr}_{\mathbf{PAKT}}(x \dot{\rightarrow}(y \dot{\rightarrow} x \dot{\wedge} y))$, the last clause in $\phi(x,X)_{mc^*}$ will give us $(y \dot{\rightarrow} x \dot{\wedge} y) \in \Phi_{mc^*}(X)$. One more application of the operator, with the same reasoning, results in $(x \dot{\wedge} y) \in \Phi_{mc^*}(\Phi_{mc^*}(X))$. That is, the inclusion of the conjunction happens one stage later. In other words, $\Phi_{mc^*}(X) \neq \Phi_{mc}(X)$, so the inductive definitions cannot be the same.

However, as we have stressed, the fixed points are indeed the same. A reasoning akin to the one just given can be used to show that all sentences that are included in the fixed point of Φ_{mc} obtained from a given initial set Y will eventually be included in the fixed point of Φ_{mc^*} with the same initial set. This happens even for the last two clauses in $\phi_{mc}(x,X)$, if we take the formalization of 'The extension of T is a maximal consistent set of sentences of $\mathcal{L}_1^+$' to be something like $\forall x(\mathrm{Sent}_{\mathcal{L}_1^+}(x) \rightarrow \neg\mathrm{T}(x) \leftrightarrow \mathrm{T}(\dot{\neg}x))$. Since the code of this formula is in the extension of $D(x)$, one can derive $\neg\mathrm{T}(x) \leftrightarrow \mathrm{T}(\dot{\neg}x)$ for each $x \in \mathrm{Sent}_{\mathcal{L}_1^+}$ through the clause for provability in **PAKT**. Then, one can obtain $\ulcorner\mathrm{T}(s) \vee \mathrm{T}(\dot{\neg}t)\urcorner \in X$ when $\mathtt{val}(t) = \mathtt{val}(s)$) (as long as we consider **PAKT** to include axioms for identity), and the same with $\ulcorner\neg(\mathrm{T}(s) \wedge \mathrm{T}(\dot{\neg}t))\urcorner \in X$.

This being said, and despite the identity of all of their fixed points, we will be working with the first proposed inductive definition— $\phi(x, X)_{mc}$ — instead of the second one, for it will make the comparison with Stern's theory much easier. Now, we claim:

Lemma 3. ***MG2****, i.e.* $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \mid\!\frac{\ \ }{\mathrm{D}}\ \varphi\}$*, is* $I_{\Phi_{mc}}$*, the minimal fixed point of the operator* Φ_{mc}*.*

Proof. This is done in two steps: we first show that **MG2** is contained in the minimal fixed point, and then we show that it is a fixed point of the operator Φ_{mc}, which will entail that **MG2** is indeed the minimal fixed point.

Thus, we start by proving that $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \subseteq I_{\Phi_{mc}}$ by induction on the length of proofs in $\mathbb{N}$-logic (i.e. provability as captured by the relation $\mid\!\frac{\ \ }{\mathrm{D}}$). For the base case (when the length is 1, i.e. when $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \mid\!\frac{\ \ }{\mathrm{D}}\ \varphi$ for φ an axiom of logic, a member of $\Delta\ulcorner\xi\urcorner$, or an atomic or negated atomic sentence true in $\mathbb{N}$), note that:

(i) the first disjunct in square brackets—i.e. the clause that includes the predicate $D(x)$—will give us all members of $\Delta\ulcorner\xi\urcorner$;

(ii) **PAKT** $\vdash \varphi$ for φ an axiom of logic or a true atomic/negated atomic sentence, and so, when $x = \ulcorner\varphi\urcorner$, letting $y \in \Delta\ulcorner\xi\urcorner$, we get

$$\mathbb{N} \models \exists y(y \in \Phi_{mc}(\varnothing) \land \mathrm{Pr}_{\mathbf{PAKT}}(y \dot{\rightarrow} x)),$$

whence $x \in I_{\Phi_{mc}}$.

For the inductive case we check MP and the ω-rule. The latter is covered by the clause for the universal quantifier. For the former, the IH gives $\ulcorner\varphi \rightarrow \psi\urcorner \in I_{\Phi_{mc}}$ and $\ulcorner\varphi\urcorner \in I_{\Phi_{mc}}$. By the clause for conjunction, we get $\ulcorner(\varphi \rightarrow \psi) \land \varphi\urcorner \in I_{\Phi_{mc}}$. Then we just need to note that **PAKT** $\vdash ((\varphi \rightarrow \psi) \land \varphi) \rightarrow \psi$, and the closure of $I_{\Phi_{mc}}$ under provability in **PAKT** does the rest.

The second step is to check that **MG2** is indeed a fixed point of the operator Φ, that is,

$$\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \mid\!\frac{\ \ }{\mathrm{D}}\ \varphi\} = \Phi_{mc}(\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \mid\!\frac{\ \ }{\mathrm{D}}\ \varphi\})$$

For the left-to-right direction, it is easy to see that, for any set X, $\Phi_{mc}(X) \supseteq X$. For the right-to-left: this is immediate if $x \in \Phi_{mc}(\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \mid\!\frac{\ \ }{\mathrm{D}}\ \varphi\})$ in virtue of the first disjunct, and almost immediate when $x \in \Phi_{mc}(\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \mid\!\frac{\ \ }{\mathrm{D}}\ \varphi\})$ in virtue of the next four disjuncts. The peculiar case is that of the universal quantifier: we know that we have a proof of each $\varphi t/v$ for all t, and we need to appeal to Zermelo's well-ordering theorem to pile the proofs together and be able to apply the ω-rule. For the case in which $x = \ulcorner \mathrm{T}(t) \urcorner$, with $t = \ulcorner \psi \urcorner$, and we have $\psi \in \{\varphi | (\mathbb{N}, \Delta^{\ulcorner}\xi^{\urcorner}) \mid\!\frac{}{\mathrm{D}} \varphi\}$, the claim follows by material adequacy. The same goes for $x = \ulcorner \neg \mathrm{T}(t) \urcorner$. For the third-to-last disjunct, we note that PAKT $\vdash \varphi \rightarrow \psi$ implies $(\mathbb{N}, \Delta^{\ulcorner}\xi^{\urcorner}) \mid\!\frac{}{\mathrm{D}} \varphi \rightarrow \psi$. Finally, we have the cases $x = \ulcorner \mathrm{T}(n) \vee \mathrm{T}(\underset{\cdot}{\neg} m) \urcorner$ and $x = \ulcorner \neg(\mathrm{T}(n) \wedge \mathrm{T}(\underset{\cdot}{\neg} m)) \urcorner$, for $n = m$. Now, if $\mathtt{val}(n) = \mathtt{val}(m)$), then $(\mathbb{N}, \Delta^{\ulcorner}\xi^{\urcorner}) \mid\!\frac{}{\mathrm{D}}$ $n = m$. The claim then follows from the observation that $(\mathbb{N}, \Delta^{\ulcorner}\xi^{\urcorner}) \mid\!\frac{}{\mathrm{D}} \mathrm{T}(t) \vee \mathrm{T}(\underset{\cdot}{\neg} t)$ and $(\mathbb{N}, \Delta^{\ulcorner}\xi^{\urcorner}) \mid\!\frac{}{\mathrm{D}} \neg(\mathrm{T}(t) \wedge \mathrm{T}(\underset{\cdot}{\neg} t))$ for any $t \in \mathrm{Sent}_{\mathcal{L}_1^+}$ (by Theorem 14).

□

This means that we can think of **MG2** as the minimal fixed point of the inductive definition *à la* McGee, but also as the minimal fixed point of the arithmetical operator *à la* Stern; put otherwise, both minimal fixed points coincide.

We can refine this result: as it happens, all ω-consistent fixed points of the two inductive definitions coincide. In order to prove this, we need to introduce the following definition, which expands on Definition 3:

Definition 4. *Given the standard model of arithmetic* $\mathbb{N}$*, a partial interpretation* $(\mathbb{N}, \Gamma)$*, and a set of sentences* X*, we write* $(\mathbb{N}, \Gamma) \mid\!\frac{X}{\mathrm{D}} \varphi$ *(read:* φ *is definitely true in* $(\mathbb{N}, \Gamma)$ *under* X *in the proof-theoretic sense) iff there is a derivation sequence that ends with* φ *and whose only members are:*

i) a member of Γ.

ii) an atomic or negated atomic sentence true in $\mathbb{N}$.

iii) an axiom of first-order logic.

iv) a member of the set X.

v) the result of applying modus ponens to previous members of the sequence.

vi) the result of applying the ω*-rule to previous members of the sequence.*

Clearly, the relation $(\mathcal{M}, \Gamma) \mid\!\frac{}{\mathrm{D}} \varphi$ presented in Definition 3 is just a special case of the relation $(\mathcal{M}, \Gamma) \mid\!\frac{X}{\mathrm{D}} \varphi$, namely, the case in which $X = \varnothing$ (i.e., $(\mathcal{M}, \Gamma) \mid\!\frac{\varnothing}{\mathrm{D}} \varphi$).

Now, consider once again the inductive definition

$$
\begin{aligned}
\forall x[R(x) \leftrightarrow \{x \in Sent_{\mathcal{L}_1^+} \wedge \\
&[G(x) \\
&\vee x \text{ is an axiom of logic} \\
&\vee x \in \text{True}_0 \\
&\vee \exists y, v(x = (\dot{\forall} v y) \wedge \forall z R(y\, {}^{\dot{z}}/_{v})) \\
&\vee \exists y(R(y) \wedge R(y \dot{\rightarrow} x))]\}]
\end{aligned}
$$

We will write $\text{FP}^X_{R,G}$ for the fixed point of the displayed inductive definition obtained when we run the fixed-point construction over X—that is, when, in the construction of the fixed point, we let X be the extension of $R(x)$ at the initial stage—and when we use the formula $G(x)$ in the definition. Accordingly, letting $D(x)$ be the formula whose extension is the set of codes of sentences in $\Delta\ulcorner\xi\urcorner$ (as understood above), $\text{FP}^X_{R,D}$ is the corresponding fixed point over X for the inductive definition with $G(x)$ substituted for $D(x)$. If, analogously, we let $\text{FP}^X_{\Phi_{mc}}$ be the fixed point obtained from the operator Φ_{mc} that inputs X at the initial stage, the result we are after is the following:

Theorem 16. *Let $X \cup \{\varphi | \ulcorner\varphi\urcorner \in \text{True}_0\} \cup \Delta\ulcorner\xi\urcorner$ be ω-consistent. Then $FP^X_{\Phi_{mc}} = FP^X_{R,D}$.*

For this theorem, we first need a couple of previous results. The strategy is simple: we show how $\text{FP}^X_{R.D}$ is equal to the equivalent of the theory **MG2** under the relation $\mid\frac{X}{\text{D}}$—that is, $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \mid\frac{X}{\text{D}} \varphi\}$—and then show that the latter is equal to $\text{FP}^X_{\Phi_{mc}}$.

Lemma 4. *Let $(\mathbb{N}, \Gamma)$ be a partial interpretation. Let $X \cup \{\varphi | \ulcorner\varphi\urcorner \in \text{True}_0\} \cup \Gamma$ be ω-consistent, and let $G(x)$ be a formula whose extension is the set of codes of sentences in Γ. Then:*

$$FP^X_{R,G} = \{\varphi | (\mathbb{N}, \Gamma) \mid\frac{X}{\text{D}} \varphi\}$$

Proof. The right-to-left direction is a routine induction on the length of the proofs under the $\mid\frac{X}{\text{D}}$ relation. The only peculiar feature is when, in proving the base case, we have a one-step proof of a member of the set X. Since the stages of the construction of $\text{FP}^X_{R,G}$ yield non-strictly increasing sets of sentences, and since all $x \in X$ were in the extension of $R(x)$ at the initial stage, they will also be in the fixed point.

The left-to-right direction is also straightforward: if $x \in \text{FP}^X_{R,G}$ in virtue of being a member of X, or of the extension of G, or of True_0, or in virtue of being an axiom of logic, then there is a one-step proof of it in $\big|\!\frac{X}{\mathrm{D}}$. The other two disjuncts of the inductive definition are covered by the ω-rule (in which case, just like in Lemma 3, we appeal to Zermelo's well-ordering theorem), and modus ponens.

$\square$

Lemma 5. *Let* $X \cup \{\varphi | \ulcorner\varphi\urcorner \in \mathrm{True}_0\} \cup \Delta\ulcorner\xi\urcorner$ *be ω-consistent. The theory* $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \varphi\}$ *is a conservative extension of* $\mathbf{PA}_\omega \cup X$ *and* $\Delta\ulcorner\xi\urcorner$ *is a materially adequate theory of truth for* $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \varphi\}$.[17]

We do not offer a proof of this lemma, for it is an exact reproduction of the (rather long) proof in [McG91, Th. 8.13]. Instead, we note two things. First, that we imposed that $X \cup \{\varphi | \ulcorner\varphi\urcorner \in True_0\} \cup \Delta\ulcorner\xi\urcorner$ be ω-consistent because, otherwise: (i) clearly, the theory would not be a conservative extension of $\mathbf{PA}_\omega \cup X$, and (ii) we would not be able to apply McGee's result, as this requires that $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \varphi\}$ and $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \neg\varphi\}$ be disjoint sets. The second thing to note is that a fundamental requisite of Theorem 8.13 (namely, that $E_\infty \subseteq \{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \varphi\}$ and $A_\infty \subseteq \{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \neg\varphi\}$) is met,[18] insofar as the axioms of **KF** are included in $\Delta\ulcorner\xi\urcorner$. These observations are what we need in order to know that Lemma 5 can be obtained following McGee's proof.

The last lemma we need in order to prove Theorem 16 is the following:

Lemma 6. $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \varphi\} = FP^X_{\Phi_{mc}}$.

Proof. Just run the proof of Lemma 3 with the new relation $\big|\!\frac{X}{\mathrm{D}}$. Note that, for the left-to-right direction, when proving the base case, we encounter $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \varphi$, for $\ulcorner\varphi\urcorner$ a member of X. For this, we just know that $\mathbf{PAKT} \vdash \psi \to \psi$ for any $\psi \in \mathrm{Sent}_{\mathcal{L}_1^+}$, so the fact that $\ulcorner\varphi\urcorner = x \in X$ yields

$$\mathbb{N} \vDash \exists y (y \in X \wedge \mathrm{Pr}_{\mathbf{PAKT}}(y \dot{\to} x)).$$

Hence, $\ulcorner\varphi\urcorner \in \Phi_{mc}(X) \subseteq \mathrm{FP}^X_{\Phi_{mc}}$.

Moreover, when tackling the right-to-left direction, Lemma 5, by proving that we have a materially adequate truth predicate in the theory $\{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\frac{X}{\mathrm{D}}\, \varphi\}$, guarantees that the clauses for the truth predicate in $\phi_{mc}(x, X)$ can be mimicked in the theory through material adequacy. That is: if

$$x = (\dot{\mathrm{T}}t), \text{ and } \mathtt{val}(t) = \ulcorner\psi\urcorner, \text{ and } \psi \in \{\varphi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \big|\!\tfrac{X}{\mathrm{D}}\, \varphi\}$$

[17] $\mathbf{PA}_\omega$ is the theory **PA** equipped with the ω-rule.

[18] Here, E_∞ and A_∞ are, respectively, the extension and the antiextension of the truth predicate in Kripke's SK minimal fixed-point construction.

then the material adequacy of T for the theory $\{\varphi|(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models^{X}_{\mathrm{D}} \varphi\}$ implies

$$(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models^{X}_{\mathrm{D}} \mathrm{T}(t).$$

□

Proof of Theorem 16. Immediate by Lemma 4 and Lemma 6. □

2.5.2 Relation between MG2 and SSK$_{mc}$

In this subsection, we show that the minimal fixed point of one of Stern's semantic theories is a subset of McGee's **MG2**.

Consider again the language $\mathcal{L}_0^+$, that is, the language of arithmetic with the truth predicate T. Let I_{SSK}, I_{SSK_c} and $I_{SSK_{mc}}$ be as in section 3, i.e., the minimal fixed points of the SSK, SSK$_c$ and SSK$_{mc}$ operators; they are all formulated in the language $\mathcal{L}_0^+$. On the other hand, the theory **MG2** is formulated in the language $\mathcal{L}_1^+ := \mathcal{L}_0^+ \cup \{\mathrm{Kr}\}$. Therefore, $\mathcal{L}_0^+$ is a sublanguage of $\mathcal{L}_1^+$. As a result of that, we need not perform a translation between the languages when showing how the minimal fixed point of one of Stern's theories is included in **MG2**.

Now, the results we obtained in subsection 2.5.1 help us prove the main result of this subsection:

Lemma 7. $I_{SSK_{mc}} \subseteq \{\varphi|(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\mathrm{D}} \varphi\}$

Proof. Thanks to theorem 3.5 in Stern, we just need to prove that $\ulcorner\varphi\urcorner \in I_{\Theta_{mc}} \Rightarrow \ulcorner\varphi\urcorner \in \{\psi|(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\mathrm{D}} \psi\}$. Θ_{mc} is the operator Stern defines in [Ste18, 831]. It is just like Θ_c, presented in section 2.3, except that, instead of $\xi(x, X) \vee \mathrm{con}(x)$, we write $\xi(x, X) \vee \mathrm{con}(x) \vee \mathrm{com}(x)$, where $\mathrm{com}(x)$ is defined as $\exists s, t(x = \ulcorner\mathrm{T}(s) \vee \mathrm{T}(\dot{\neg} t)\urcorner \wedge \mathtt{val}(s) = \mathtt{val}(t))$. We can call this whole formula $\xi_{mc}(x, X)$. We thus investigate all possible cases in which a sentence φ can be a member of $I_{SSK_{mc}}$, based on the operator Θ_{mc}. Let $a = \ulcorner\varphi\urcorner$.

- When φ is a true atomic or negated atomic sentence of the language $\mathcal{L}_0$, then clearly $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\mathrm{D}} \varphi$.

- When $a = \ulcorner\varphi\urcorner \in I_{\Theta_{mc}}$ in virtue of $\mathrm{con}(a)$,

$$\exists s, t(a = \ulcorner\neg(\mathrm{T}(s) \wedge \mathrm{T}(\dot{\neg} t))\urcorner \wedge \mathtt{val}(s) = \mathtt{val}(t)),$$

we can reason in two different ways. Since the same applies to the case $\mathrm{com}(a)$, we expound one way for this case, and leave the other form of reasoning for the other. Thus, we appeal here to the principles of T in $\mathbb{N}, \Delta\ulcorner\xi\urcorner$. Assuming $s = n$, $t = m$, we have $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\overline{\mathrm{D}}} n = m$, and also $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\overline{\mathrm{D}}} \neg(\mathrm{T}(n) \wedge \mathrm{T}(\dot{\neg} n))$ by CONS, so $(\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\overline{\mathrm{D}}} \neg(\mathrm{T}(n) \wedge \mathrm{T}(\dot{\neg} m))$ by applying logic inside the truth predicate (remember Lemma 2 and principle (*) in the proof of Theorem 14).

- Suppose $a \in I_{\Theta_{mc}}$ in virtue of $\mathrm{com}(x)$, that is,

$$\exists s, t(a = \ulcorner\mathrm{T}(s) \vee \mathrm{T}(\dot{\neg} t))\urcorner \wedge \mathtt{val}(s) = \mathtt{val}(t)).$$

  As mentioned, we could have applied the same reasoning as in the case of $\mathrm{con}(a)$, but there is an alternative way to prove it. We just appeal to the second-to-last clause in $\phi_{mc}(x, X)$, so that if $\ulcorner\varphi\urcorner = a \in I_{\Theta_{mc}}$ because $\mathrm{com}(a)$, then $\phi_{mc}(a, S)$ for any set S, including the empty set. So $a \in \{\psi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\overline{\mathrm{D}}} \psi\}$.

- When $a \in I_{\Theta_{mc}}$ due to one of the clauses (2.2)-(2.7) in the definition of $\xi(x, X)$ (that is, the clauses for the connectives and for the truth predicate—see section 2.3 above), just note that the clauses are identical to the clauses in $\phi_{mc}(x, X)$, which means that $a \in I_{\Phi_{mc}} = \{\psi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \vdash \psi\}$.

- The final case is the case

$$\mathbb{N} \models \exists y(\xi_{mc}(y, S) \wedge \mathrm{Pr}_{\mathbf{PAT}}(y \dot{\rightarrow} a)).$$

  Then, there exists some $n \in \omega$ such that $\mathbb{N} \models \xi_{mc}(\bar{n}, S) \wedge \mathrm{Pr}_{\mathbf{PAT}}(n \dot{\rightarrow} a)$. Just like $a = \ulcorner\varphi\urcorner$, let $n = \ulcorner\chi\urcorner$. The above shows that $\mathbb{N} \models \xi_{mc}(\bar{n}, S)$ implies

$$\chi \in \{\psi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\overline{\mathrm{D}}} \psi\}.$$

  Moreover, $\mathbf{PAT} \vdash \chi \rightarrow \phi$ implies $\mathbf{PAKT} \vdash \chi \rightarrow \phi$. Write $\Phi^{\alpha}_{mc}(\varnothing)$ for the α-th stage in the construction of the minimal fixed point of Φ_{mc}. Then, since

$$\chi \in \{\psi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\overline{\mathrm{D}}} \psi\}$$

  entails that there is some ordinal α such that $\ulcorner\chi\urcorner = n \in \Phi^{\alpha}_{mc}(\varnothing)$, we have

$$\mathbb{N} \vDash \mathrm{Sent}_{\mathcal{L}_1^+}(a) \wedge \exists x(x \in \Phi^{\alpha}_{mc}(\varnothing) \wedge \mathrm{Pr}_{\mathbf{PAKT}}(x \dot{\rightarrow} a)).$$

From this, it follows that $\mathbb{N} \vDash \phi_{mc}(a, \Phi^{\alpha}_{mc}(\varnothing))$, and we can conclude $a \in \Phi^{\alpha+1}_{mc}(\varnothing) \subseteq \{\psi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\mathrm{D}} \psi\}$.

□

This gives us the inclusion of the minimal fixed point of $\mathbf{SSK}_{mc}$ in **MG2**. We know that the reverse direction does not hold, at least not straightforwardly. This is due to the fact that, as we mentioned, $\mathcal{L}_0^+$ is a sublanguage of $\mathcal{L}_1^+$, which ensures that there will be sentences proved by **MG2** that cannot be in $I_{SSK_{mc}}$ for mere linguistic reasons—in particular, all the sentences that include the auxiliary truth predicate Kr.

However, it is **still open** whether the reversed inclusion can hold modulo an appropriate translation between the two languages. That is, the question is whether there is a natural translation function $\rho : \mathcal{L}_1^+ \mapsto \mathcal{L}_0^+$ such that

$$\phi \in \{\psi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\mathrm{D}} \psi\} \Rightarrow \ulcorner\rho(\phi)\urcorner \in I_{SSK_{mc}}$$

for all $\ulcorner\phi\urcorner \in \mathrm{Sent}_{\mathcal{L}_1^+}$.

In any case, we note a corollary that follows from Lemma 7. $I_{SSK_{mc}}$ is equal to I_{mc}, the minimal fixed point of the supervaluation scheme mc.[19] As a result:

Corollary 2. $I_{mc} \subseteq \{\psi | (\mathbb{N}, \Delta\ulcorner\xi\urcorner) \models_{\mathrm{D}} \psi\}$.

2.6 What comes next

In this chapter, we presented three different ways of reaching what we have called supervaluational truth: Kripke-van Fraassian fixed-point semantics, Stern's supervaluational-style truth, and McGee's materially adequate theory **MG2**. We have presented and/or established some relations between them, the most novel perhaps being the fact that McGee's theory is a superset of the minimal fixed point of Stern's theory ($\mathbf{SSK}_{mc}$), as well of the minimal fixed point obtained with the traditional supervaluation scheme mc. It is hoped that this results will shed light when it comes to constructing our own theory of determinate truth.

To close the chapter, we list down some questions that either were left open during the chapter or can be posed now on the basis of its content, and whose answer may continue to provide insight into how supervaluational truth works:

[19] This is stated without proof in [Ste18, 832]

- What is the proof-theoretic strength of the theory **VFM** (cf. section 2.2)?
- Do all fixed points of Meadow's double-conditional dependent truth operator coincide with the fixed points of the supervaluational scheme vb (cf. section 2.3)? And with the fixed points of Stern's **SSK**?
- Can we come up with a dependent-truth fixed-point semantics that has the same properties with respect to the supervaluation scheme vc as Meadows' has with respect to the scheme vb (cf. section 2.3)? And a dependent-truth fixed-point semantics that has these properties with respect to the scheme mc?
- What are the principles of the truth predicate of McGee's **MG1** and how does this theory relate to Stern's theories **SSK** and $\mathbf{SSK}_c$ (cf. sections 2.4 and 2.5)?
- Are the minimal fixed point of McGee's theory and the minimal fixed point of $\mathbf{SSK}_{mc}$ the same, modulo a natural translation from $\mathcal{L}_1^+$ to $\mathcal{L}_0^+$? If so, are all the fixed points of the inductive definitions defined in section 2.5 equal to the fixed points of either the supervaluation scheme mc, or $\mathbf{SSK}_{mc}$?

Chapter 3

Chapter 2 constituted a first step towards building a theory of truth and determinate truth. The task is still very much unfinished: we have just outlined key points of supervaluation-like treatments of truth on which, we hope, our theory of determinate truth will be built. We intend to complete that part of the project in incoming research. In the meantime, we go back to the theme of Chapter 1: *arithmetical* determinacy. This chapter will explore a way of securing **arithmetical determinacy**, that is, the thesis that every arithmetical statement has a determinate truth-value. And we do it by drawing on the so-called Isaacson's thesis, proposed by Daniel Isaacson [Isa87, Isa92]. According to this thesis, being arithmetical *just is* being provable or refutable in Peano Arithmetic. So, if the thesis is sound, arithmetical determinacy is guaranteed: we need only look at whether **PA** proves or refutes a statement to know its determinate truth-value. And, if **PA** neither proves nor refutes the statement, then we also know that the statement is not arithmetical, and hence irrelevant for **arithmetical determinacy**.

The challenge is, of course, to show that Isaacson's thesis is an attractive and sound one. A few steps in this line have already been taken besides the work of Isaacson—e.g. [Smi08, TB18]. In this chapter, we work on that direction too. In particular, we identify an important challenge for Isaacson's thesis, and argue that the thesis can meet it. Thus, we propose a reading of Isaacson's work in which the status of certain **PA**-provable sentences as arithmetical, at least in the sense of the word Isaacson proposes, can be called into question. We first introduce a notion of *adequacy* for arithmetical theories that refers back to the set of arithmetical statements that follows from Isaacson's conception of arithmetical truth, and note that adequacy thus understood is a desirable feature of arithmetical theories. We then argue that, under the aforementioned reading, **PA** seems to be inadequate with respect to arithmetical truth—that is, some of the truths proven by **PA** are not arithmetical truths in the sense of Isaacson. We then try to show that the way in which Isaacson, who had foreseen the reading that leads to the inadequacy claim, tries to prevent the latter, is not entirely satisfactory. Finally, we explore a different route to restore adequacy: justifying the arithmetical nature, in Isaacson's sense, of those claims that motivated the inadequacy concern in the first place. As a paradigmatic case study, we try to understand how it can be that **PA**, the theory of finite mathematics in Isaacson's conception, can prove instances of transfinite induction for ordinals well beyond ω.

3.1 Isaacson's thesis

Ever since at least Tarski, the mainstream conception of arithmetical truth has equated the latter with satisfiability in the standard model for the language of arithmetic, that we shall call $\mathcal{L}_0$ and which includes the nonlogical constants $(\mathrm{S}, 0, +, \cdot, <)$.[1] We refer to this model simply as the standard model of arithmetic $\mathcal{N}$, and to the set of sentences true in this model as true arithmetic, or Th($\mathcal{N}$)[BBJ07, 295].

Contra this widespread view on arithmetical truth, Daniel Isaacson holds that the notion of being arithmetical is not only formal but also *epistemic*, since '[i]t has to do with the way in which we are able to perceive [a] statement's truth or falsity' [Isa92, 95]. Isaacson's understanding of arithmetical truth, then, comes in the form of a recursive definition. The base clause asserts that a true statement is arithmetical when its truth can be seen to follow directly from our understanding of the natural number structure; he seems to think that the axioms of **PA** (and perhaps those alone) are arithmetical in this sense. The recursive clause asserts that a true statement is arithmetical if its truth can be perceived as such through first-order logical inferences from known truths whose arithmetical nature has been granted.[2] Thus:

> [A] truth expressed in the language of arithmetic is arithmetical just in case its truth is directly perceivable so expressed, or on the basis of other

[1] For differing views, see [Say90].

[2] One can question, in any case, the appropriateness of this recursive definition. The problem has to do with the base clause: while **PA** undeniably captures the first-order content of second-order arithmetic, so do the different theories that are mutually elementary reducible with standard **PA**, such as $\bigcup_n \mathrm{I}\Sigma_n$ (for a definition of elementary reducibility see, e.g., [NLeyng]). It seems hence arbitrary to establish that one set of axioms, and not the other, can be directly perceived as the set of truths about the natural number structure. They all correspond to different axiomatizations of what we consider first-order arithmetic with full induction to be. In sum, the base clause ought to allow for a wider range of applicability regarding what counts as following directly from our understanding of the concept of natural number.

> truths in the language of arithmetic which are themselves arithmetical. The analysis of the number concept as discussed in §§ 2, 3, 4, seems to me to render the axioms of Peano Arithmetic arithmetical, in the sense that their truth is directly perceivable so expressed, and on this basis the second clause renders the theorems of PA arithmetical. [Isa87, 162],

Admittedly, Isaacson's recursive definition only accounts for arithmetical truths, that is, 'being arithmetical' is a property that applies only to certain statements (at least on paper, for Isaacson sometimes speaks of arithmetical concepts, as implying that concepts can also be arithmetical; more on this later). But one can easily account for arithmetical falsities by stipulating that all statements the negation of which is arithmetical can be regarded as arithmetical too.[3] Non-arithmetical statements, on the contrary, are those incorporating what he calls 'higher-order notions': syntactic concepts such as 'consistency' or 'provability', which are not implicit in our understanding of the natural numbers; but also infinitary notions, 'in the sense of presupposing an infinite totality' in the words of Isaacson [Isa87, 155], as opposed to finitary notions. Accordingly, non-arithmetical statements will be known as higher-order statements. Plus, to be clear, from now on, and unless otherwise specified, when we speak of 'arithmetical' we will mean 'arithmetical in Isaacson's sense'; and, when we speak of 'arithmeticality', we will mean 'the status of being arithmetical in Isaacson's sense'.

Our current understanding of the natural number structure owes much to Dedekind's and Frege's studies of the principles of arithmetic, and so theirs (and perhaps Dedekind's to a greater extent) are seen as the best categorical conceptual analysis of the notion of natural number. Admittedly, Dedekind's analysis contains higher-order concepts in the form of second-order quantification over subsets of natural numbers. But what remains when we strip this analysis of its second-order content—i.e., when we 'first-orderize' this second-order quantification—is just **PA**. As a result, Isaacson goes, **PA** enjoys a privileged position among all axiomatizations of elementary number theory: not only does the analysis of the natural number structure allow us to perceive **PA** as true and strictly arithmetical, but it is also the case that **PA** captures all there is to *arithmetical*—as opposed to just *mathematical*—truth: if a statement in the language of arithmetic is not provable in **PA**, then some 'hidden' higher-order concept is needed either to directly perceive its truth or the proof of it.

With this in mind, we offer a precise formulation of Isaacson's thesis. There are

[3] As he made clear to me, this is Isaacson's preferred way to account for arithmetical falsities.

a couple of different phrasings in the literature (see e.g. [Hor01, Smi13]). Isaacson's seemingly preferred way to put it is that Peano Arithmetic consists of those truths which can be perceived as truths either directly or via a proof from the purely arithmetical content of the categorical conceptual analysis of the notion of natural number. However, and since we already know that 'those truths which...' is just short for Isaacson's notion of arithmetical truth, we follow Luca Incurvati's [Inc08] shorter wording:

> ***Isaacson's thesis*** Peano Arithmetic is *sound* and *complete* with respect to arithmetical truth (in the sense of Isaacson).

As we see it, Isaacson's thesis gains a great deal of plausibility from the fact that it captures the long-standing mathematical intuition that our natural number system is at the heart of all finite mathematics, and that **PA** is the set of axioms that best captures such a system (or even: that **PA** is the axiomatization of such a system). Even so, the thesis must be tested, and its most pressing challenge the consists in accommodating the kind of sentences which traditionally rendered support for the belief that **PA** is incomplete: sentences expressible in $\mathcal{L}_0$ and satisfied by $\mathcal{N}$ but independent of **PA**. The thesis projects that all these sentences present a common feature, namely their not being arithmetical in nature. Two clear examples Isaacson examines are the Gödel sentence for **PA** and Goodstein's theorem. In the first case, the arithmeticality of the sentence is denied on the basis that seeing its truth requires understanding the notion of provability in **PA**, the kind of syntactic higher-order notion that does not follow from our grasp of the natural number structure. As per Goodstein's theorem, the proof of the theorem relies on the well-ordering of ordinals (i.e. transfinite induction) up to ε_0 ($\mathtt{TI}(\varepsilon_0)$ henceforth). The latter, however, is known to entail, over **PA**, the sentence Con(**PA**) (i.e., the sentence asserting the consistency of **PA**), and hence is also higher-order in nature. As a result, we should not expect **PA** to prove neither the Gödel sentence nor Goodstein's theorem, so Isaacson's thesis stands.

Similar reasonings are given for two further well-known theorems independent of **PA**: the Paris-Harrington theorem and Friedman's finitization of Kruskal's theorem. Thus, although none of these arguments is conclusive enough to secure Isaacson's thesis—what happens, for instance, with the Kanamori-McAloon theorem or **PA**-unprovable versions of the graph minor theorem [Bov09]?—they make it rather convincing. In other words, they seem to indicate that all arithmetically-expressible theorems that **PA** cannot prove aren't, after all, arithmetical truths.

3.2 Relevance for arithmetical determinacy

Once it has been presented, let us say a bit more on why we are interested in Isaacson's thesis. As we outlined in the introduction to this chapter, it all has to do with providing the basis for **arithmetical determinacy**, the assertion that all arithmetical statements have a determinate truth-value.

Following the previous section, Isaacson's thesis amounts to the idea that **PA** is *complete* with respect to arithmetical truth—that is, for any *arithmetical* statement φ, either $\mathbf{PA} \vdash \varphi$ or $\mathbf{PA} \vdash \neg\varphi$. Once again, *arithmetical* is here understood in Isaacson's restricted sense. This being the case, under the reasonable assumption that provability in **PA** suffices to establish the truth of a statement, and refutability to establish its falsity,[4] and on the assumption that **PA** is consistent, then all arithmetical statements have but one truth-value. Of this one truth-value we can say that is the *determinate* truth-value of the statement. Therefore, under Isaacson's thesis we have a straightforward, affirmative answer to the question 'Do all arithmetical statement have a determinate truth-value?'.

Now, Chapter 1 presented what we then called the metasemantic challenge for mathematical determinacy, raised by Warren and Waxman [WW20a]. It could be summed up as the idea that mathematical determinacy, if it arises, needs to be explained. We also indicated that the challenge transfers word by word to the case of arithmetical determinacy. But now, thanks to Isaacson's thesis, it seems that we can meet the metasemantic challenge for arithmetical determinacy—or, at least, partially. Indeed, Isaacson's thesis establishes that arithmetical determinacy arises (i.e., that **arithmetical determinacy**, in bold, is true). Moreover, and to a certain extent, it explains *why* it arises: because the limits of what counts as arithmetic are determined jointly by formal and epistemic constraints, and **PA** alone serves to decide all claims within those limits.

This proposal may generate some reticence, though. First, one may be worried that Isaacson's thesis still walks on thin ice. The remaining of this chapter aims to bring the reader closer to conviction of the opposite. Secondly, one could expect a more fine-grained account of the arising of arithmetical determinacy. For instance, should one believe that set theory is indeterminate, one may demand an explanation of why something like Isaacson's thesis cannot be used to secure the determinacy of set theory.[5] Wherein lies the difference between set theory and arithmetic

[4] An assumption that follows, of course, from the soundness of **PA**.

[5] Note that Leon Horsten [Hor01] has, in fact, defended an analogue of Isaacson's thesis for set theory—in particular, for **ZFC**.

that renders one indeterminate and one determinate? An exploration along these lines will also be the future object of study of this project.

A third point: one may advance Putnam-style worries against Isaacson's thesis in the first place. After all, the thesis draws on concepts like 'finite', 'natural number', and so on, which Putnam may have shown to be problematic. I take these worries to be well-founded, as chapter 1 demonstrated; and I think that much could be said on whether endorsing Isaacson's thesis resolves the Putnamian tensions, but we will not do that here. Yet a brief, relevant response may run as follows: Isaacson's thesis teaches us to shift the focus from a model-theoretic to a first-order axiomatic conception of truth, i.e., truth as being exclusively what the axioms, equipped with first-order logic, can capture. It seems to me that this dissolves the strength of the model-theoretic arguments, which precisely rely on the variety of models for the same first-order arithmetical theory. Whether this is enough, or correct at all in the first place, will not be answered here for reasons of space.

In sum: for now, our modest goal is to provide evidence that speaks in favor of Isaacson's thesis, as a way to make some progress towards meeting the metasemantic challenge. As we mentioned, the way we pursue this goal is by identifying a problem that represents an important obstacle to the plausibility of Isaacson's thesis, and showing that the thesis can respond to it appropriately. It is therefore time to expound the problem.

3.3 The inadequacy claim

One of the key points behind Isaacson's thesis is that it lifts **PA** as *the* first-order axiomatization of arithmetic, in the sense of proving all and only arithmetical truths. The 'all' part of the claim is established through completeness and it has certainly been the main focus of the literature, possibly due to its novelty after (and its defiance of) Gödel's incompleteness theorems (see [Smi08, TB18]). But the 'only' side has not been thoroughly addressed so far. This section aims to show that, under a certain reading of Isaacson's original 1987 paper, there is a real possibility of **PA** being an *inadequate* theory of arithmetic, thus motivating what we have called 'the nuanced thesis': the thesis that **PA** is complete but inadequate with respect to arithmetical truth. Here, the notion of 'adequacy' has a precise meaning, in line with Isaacson's conception of arithmetical truth, that we shall now explain.

3.3.1 The notion of adequacy

When we assert that, following Isaacson's thesis, **PA** is complete with respect to arithmetical truth, the notion of completeness differs as much from the model-theoretic notion as Isaacson's conception of arithmetical truth does from the Tarskian one. That is, we do not intend to say that **PA** proves all formulae true in the standard model of arithmetic, for this is plainly not the case. Rather, we just mean that there is no arithmetical statement in the sense of Isaacson that **PA** does not prove.

We now intend to define the counterpart of this notion, which one can understand as the analogue of soundness under Isaacson's conception of arithmetical truth. We say that an arithmetical theory T is sound iff every theorem of T is true. Under the Tarskian approach, truth just is satisfiability in the standard model of arithmetic. With our new understanding of arithmetical truth, we can define the notion of adequacy in a similar manner:

> ***Adequacy*** An arithmetical theory T is *adequate* iff every theorem of T is an *arithmetical* truth.

In other words, an arithmetical theory T is adequate iff every theorem of T is a true statement in the language of arithmetic that follows from the recursive definition proposed by Isaacson. It must be noted that, for a theory to be inadequate, it need not be unsound, that is, it need not prove a false statement in the language of arithmetic. It will suffice for it to prove a true statement in the language of arithmetic that is not true in the sense of Isaacson.

We understand that the notion of adequacy is, at the very least, desirable for an arithmetical theory. In fact, we take it to be a desirable property of most mathematical theories with a clearly defined and restricted domain (sometimes knwon as 'non-algebraic theories'), since it guarantees that they do not 'overshoot' in relation to that intended matter. To give an example, it would be rather unsettling if we were to show that, from the axioms for Euclidean geometry, one can prove the existence of a Mahlo cardinal. In the case of arithmetical theories, these are meant to describe the natural number structure. Thus, one expects an *arithmetic* theory to prove *arithmetic* results.

With this notion of adequacy in mind, we can now move on to understand in what sense we can say that **PA** is an inadequate theory under Isaacson's thesis.

3.3.2 The problem of inadequacy

We shall start by noticing, as Isaacson does, that some statements that are provable in **PA** seem to belong to the class of non-arithmetical truths—either because they are about infinitary objects, or because they are of a seemingly syntactic nature. An example of the former is transfinite induction for any ordinal $\alpha < \varepsilon_0$, that we shall denote $\mathtt{TI}(< \varepsilon_0)$;[6] clearly, ordinals like ω^{ω^3} are infinite—but **PA** shows ω^{ω^3} is well-ordered! An example of the latter is Con(**PRA**), the sentence that formalizes the consistency of Primitive Recursive Arithmetic.

But how can **PA** even speak of consistency, or of infinite ordinals? The key point, of course, is coding (broadly understood). The arithmetization of syntax allows **PA** to speak about syntactic notions, 'coding' such notions with strings of arithmetical constants; an ordinal notation system does the same in relation to infinite ordinals. And the existence of coding, Isaacson argues, suffices to realise that this kind of sentences are, after all, arithmetical in nature: as an auxiliary device, coding 'pulls the ostensibly higher-order truth into the arithmetical' [Isa87, 165] and allows for a proof of the statement in strictly arithmetical terms, which is all we need for the statement to count as arithmetical. Note that this is a consequence of Isaacson's epistemic approach to arithmetical truth: arithmeticality is not so much a feature of the statement in question but of the way we come to see its truth.

This cannot be taken, however, to be a conclusive answer, as Isaacson acknowledges. Bearing in mind the importance of the epistemic component in the definition of arithmetical truth, we could doubt whether a strictly arithmetical but immeasurably and unfollowably long proof (in the sense of it containing too many symbols) may even qualify as a way to perceive the truth of a statement at all. Especially, when considering that the introduction of higher-order notions renders the proof radically shorter. As Isaacson puts it, it seems that 'there can be cases where the higher-order perspective is essential for actual conviction as to truth of the arithmetically expressed sentence. One may know that a derivation in PA must exist, but if generated would be so long as to be unsurveyable.' [Isa87, 165]. In sum, the point is as follows: there are statements provable in **PA** but whose proofs, if they are to be the means by which we perceive the truth of the statements, require higher-order notions, as they would turn out to be too long otherwise—which means, in turn, that there are **PA**-provable statements that are not arithmetical truths. Indeed, this is, for instance, the reason why we work with infinite ordinals and not their no-

[6] The expression is a little sloppy here: $\mathtt{TI}(< \varepsilon_0)$ is a schema, that is, needs to be instantiated by some formula. Let's take that for granted in what follows.

tations in proving, e.g., $\mathtt{TI}(\omega^{\omega})$ in **PA**. Given the correctness of our ordinal notation,[7] we know that there exists a corresponding proof with formulas that strictly belong to $\mathcal{L}_0$. But such a proof would be too long to be carried out in practice, so the deployment of uncoded infinite ordinals becomes indispensable for perceiving the truth of the statement. Thus, $\mathtt{TI}(\omega^{\omega})$ becomes a statement provable in **PA** but non-arithmetical.

This situation is, according to Isaacson, one in which the notions of provability-in-practice and provability-in-principle make all the difference. Someone who accepts that provability in principle in **PA** is sufficient to define the boundaries of arithmeticality need not worry further. Insofar as a statement is in principle provable in **PA** in strictly arithmetical terms, the statement counts as arithmetical:

> If one is prepared to countenance a notion of being 'in principle' derivable in PA, then the present problem disappears. One might consider that this move is legitimate, as enabling one to define precisely a theoretical boundary, to which mathematical practice approximates. [Isa87, 166]

However, and as we have seen, Isaacson's thesis puts the emphasis on the *epistemic* character of arithmeticality. So the defender of the notion of provability-in-practice has a strong point in this context. Being arithmetical is here as much a product of our possibility to perceive the truth of the statement as it is a product of the language in which the statement can be expressed. Hence, it looks as if *followability* is a reasonable condition on what counts as a proof that allows to establish the arithmetical nature of a statement. Thus, Isaacson [Isa87, 165-66] concedes: 'I have in my discussion been considering provability in terms of providing a basis for perceiving the truth of a given statement. In these terms, a proof in PA of a given proposition being infeasibly long has to be taken seriously.'

The problem, as we have mentioned, is that the provable-in-practice attitude just presented has an important implication: that ***PA** is inadequate as an arithmetical theory*. Inadequacy here must be understood as above, namely, as implying that some statements provable in the arithmetical theory are not arithmetical. Roughly, that **PA** proves too much for a theory of arithmetic. The situation is depicted in figure 3.1 below: arithmetical truth would be a proper subset of the set of **PA**-provable truths, which is in turn a proper subset of truths expressible in $\mathcal{L}_0$ (due to Gödel's theorem).

[7] See e.g. [Poh09, Th.3.3.17] for a theorem establishing such correctness.

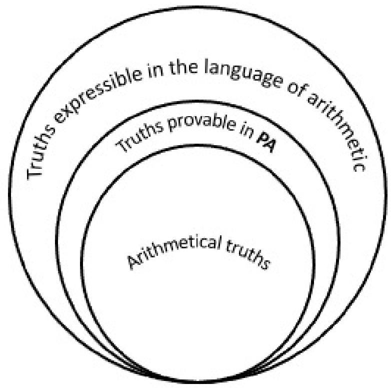

Figure 3.1: The relation between arithmetical truth, truths provable in **PA** and truths expressible in the language of arithmetic

This is something Isaacson himself acknowledges, for he grants that, should one privilege the notion of provability in practice,

> then within the arithmetically expressible truths of mathematics, we must think of the boundary between those which are purely arithmetical and those which are essentially higher-order as running somewhat inside the collection of those for which derivations in PA exist. [Isa87, 166]

In recent conversation, Isaacson has made clear to me that he favours the in-principle view. His opinion seems to be that the provability-in-practice approach puts one on the road of strict finitism, an undesirable philosophy of mathematics that Isaacson now, and unlike then, definitely rules out. Be that as it may, and as we have argued, we still think that the epistemic turn on arithmetical truth fostered by Isaacson makes a case for the in-practice reading. Thus, in the remaining of the paper we follow that reading, trying to make sense of it. In doing so, we can present a new formulation of Isaacson's thesis, to be compared with the one given before:

> ***The nuanced thesis*** Peano Arithmetic is complete and sound with respect to arithmetical truth, but *inadequate* as a theory of arithmetic.

3.3.3 Isaacson's proposed way-out

As it happens, Isaacson offers a solution to the inadequacy claim on behalf of the provability-in-practice advocate. To follow his reasoning, let us recap the problem: there are true statements in the language of arithmetic, e.g., (the coded version of) $\mathtt{TI}(\omega^{\omega})$, that can be proved either via higher-order notions embedded in a relatively short proof, or in purely arithmetical terms but with an unsurveyably long proof. Now, we could appeal to the mere existence of that proof in **PA** (even if it is humanly ungraspable) to argue that the higher-order notions are not indispensable. But, given his epistemic approach to arithmeticality, in which a proof has to be a vehicle to perceive the truth of a statement, the provability-in-practice advocate does not buy that argument. Then, and possibly with the goal of avoiding the implications linked to the inadequacy claim in mind, Isaacson makes a move on behalf of such hypothetical advocate. According to Isaacson, one could reject extremely long proofs, such as the one for $\mathtt{TI}(\omega^{\omega})$ or the one for Con(**PRA)**, as genuine proofs in **PA**. As a result, 'provable in **PA**' would acquire a new, more limited character, and the set of truths provable in **PA** would coincide with the set of arithmetical truths. This can be visualized by considering again figure 3.1: the circle that represents truths provable in **PA** 'shrinks' to the boundaries of the circle of arithmetical truths. The thesis, after all, stands.

Now, let me counter this move. There are at least two considerations as for why we might not want to reject very long proofs as genuine **PA**-proofs. First of all, doing so deprives **PA** of its privileged proof-theoretic status among first-order axiomatizations of second-order arithmetic. For suppose we formalized Isaacson's proposed notion of provability, i.e., suppose the length in symbols of the shortest **PA**-proof of, e.g., Con(**PRA**), is n, and so we only admit proofs in **PA** of length less than n. In other words, and if $\vdash_n$ is the symbol we use for this restricted notion of provability in **PA** and $\rho(x)$ is a function that gives the length in symbols of the shortest proof in **PA** of the formula represented by x, we write **PA** $\vdash_n \varphi$ iff **PA** $\vdash \varphi$ and $\rho(\varphi) < n$. The result is then that **PA** $\nvdash_n$ Con(**PRA**). Therefore, it is not clear that **PA** is in any better proof-theoretical position than, in this case, **PRA**. This, in turn, goes against Isaacson's thesis, which, after all, revolves around the privileged status of **PA** over other first-order theories of arithmetic.

In the second place, and perhaps even more relevantly for Isaacson's thesis, it seems likely that the downgrading of **PA** could happen not only at the proof-theoretic but also at the strictly number-theoretic level. In a paper on the length of proofs, Gödel [G6̈5] asserts that, for each computable function Φ, there are infinitely many different formulae x provable in **PA** (or in any first-order arithmetical theory, for that matter) such that $\rho(x) > \Phi(\rho_2(x))$, where $\rho(x)$ is defined as above and $\rho_2(x)$ is the length of the shortest proof of x in $\mathbf{PA}_2$. Now, let's suppose that, among all instances of transfinite induction up to ω^ω, the instantiation with formula φ is the one whose shortest proof involves the greatest number of symbols, and that the proof is too long to be surveyed—so that, following Isaacson's suggestion, we do not consider it as a legitimate proof for **PA**. Let's write $\rho(\ulcorner \mathtt{TI}(\omega^\omega, \varphi) \urcorner)$ for the shortest proof of the instantiation with formula φ of the transfinite induction schema up to ω^ω.[8] And, following Isaacson's suggestion, let's suppose that only proofs of length $< \rho(\ulcorner \mathtt{TI}(\omega^\omega, \varphi) \urcorner)$ are accepted. Then, we can find a computable function Ψ such that $\Psi(\rho_2(\ulcorner \mathtt{TI}(\omega^\omega, \varphi) \urcorner)) = \rho(\ulcorner \mathtt{TI}(\omega^\omega, \varphi) \urcorner)$. After that, it is not difficult to generate a countably infinite number of computable functions Ψ' that bound Ψ from above, i.e. such that

$$\Psi(n) \leq \Psi'(n), \text{ for all } n \in \mathbb{N}$$

But then, for each of those Ψ', Gödel's result tells us that there are infinitely many different formulas of $\mathcal{L}_0$ that are provable in **PA** and such that the length of their shortest proof is greater than $\Psi(\rho_2(\ulcorner \mathtt{TI}_{(}\omega^\omega, \varphi) \urcorner))$. However, all these formulas need to be considered as unprovable in **PA**, or at least as formulas the proof of which are not genuine for **PA**. There are thus infinitely many different theorems of **PA** that we stop considering as such.

If any of these two considerations seems pertinent, then provability in **PA** cannot be so freely adjusted to match the set of arithmetical truths, and we are left with the inadequacy claim under the provability-in-practice reading of Isaacson's thesis. The remaining of the paper will now be devoted to show how we can still avoid this claim with arguments different to those of Isaacson.

3.4 Resisting the nuanced thesis I

In the previous section, we argued that, according to certain reading of Isaacson's work, **PA** could be inadequate; and this is definitely a hard pill to swallow for logicians and philosophers of mathematics alike, who often take **PA** to be *the* first-order theory of arithmetic par excellence. The reading in question prioritises the notion of provability-in-practice over that of provability-in-principle. As we saw and objected to, Isaacson suggests that the provability-in-practice advocate may just do away in

[8] We write upper corners ($\ulcorner \urcorner$) to indicate that what comes inside corresponds to the 'coded', arithmetical version of the formula.

PA with all those statements the proof of which is too long to be carried out in practice. Therefore, we concluded that the provability-in-practice advocate should not take the path delineated by Isaacson. As we pointed out, it also seems that Isaacson himself would accept that conclusion now, having identified the problems of such path, which are no others than the problems faced by a strict finitist.

Nevertheless, it looks to us as if Isaacson was here conflating two views that need to be distinguished: the provability-in-practice advocate as regards arithmetical truth, and the provability-in-practice advocate as regards derivability in a theory of arithmetic. That is: one can defend that provability in practice must be a criterion for actual perceivability of the truth of a statement and thus, following Isaacson, of its arithmetical nature; and one can defend that provability in practice must be a *formal* criterion for derivability over an arithmetical theory. Only the latter seems to be equivalent to strict finitism (sometimes also known as ultrafinitism). The former, on the contrary, just concerns what we can consider arithmetical in Isaacson's sense. Now, in what follows, we try to show that the provability-in-practice as regards arithmetical truth is on safe grounds, so that even those statements that fall outside the scope of what is provable in practice can, by other means, be considered arithmetical on Isaacson's sense of the term. Actually, what we propose is more of a conjecture, namely: that there is a way to justify the arithmeticality of each of those statements—exactly the opposite of what Isaacson did with **PA**-independent statements.

How can we defend this conjecture? What follows now is a case study which has that goal in mind. We will look at two paradigmatic kinds of statements that may lead to the inadequacy problem: transfinite induction claims and consistency statements. Or rather: we will be looking at only one of these, transfinite induction claims, and, we believe, this will suffice to show that we can justify the arithmeticality of consistency claims as well. The reason why is that all claims of the form Con(T), where T is a subsystem of **PA**—that is, claims that appeal to seemingly higher-order concepts of syntactic nature—can be shown to be provable from a transfinite induction claim up to a certain ordinal below ε_0, over a subsystem of **PA** proof-theoretically weaker than T itself.[9] This follows from the fact that each of these first-order subsystems, which are weaker than **PA**, has a proof-theoretic ordinal strictly smaller than ε_0. Hence, should we show that all transfinite induction

[9] An earlier version of this chapter suggested that the reason why had to go with the entailment being provable in **PA**. But clearly this does not suffice, and I thank Giorgio Venturi for pointing this out. After all, *every* **PA**-provable statement is entailed by any other statement over **PA**. And we do not want to say that the arithmeticality of Con(**PRA**) is granted by the fact that $0 = 0$ is arithmetical. It is the special connection between transfinite induction and consistency that must do the job.

statements up to ε_0 are, after all, arithmetical, we could equally conclude that all syntactic statements of this sort are arithmetical: epistemically, the truth of the syntactic statement would be perceivable insofar as the entailment can be established in strictly arithmetical terms.

Thus, we will try to argue for our conjecture as follows. The problem of inadequacy with statements such as $\mathtt{TI}(\omega^{\omega})$ is that their not-so-long proofs make use of infinite ordinals and not their notations, which seem to be higher-order (infinitary) notions. Nonetheless, if we are able to question the very idea that these are higher-order notions and make the case for their finitary nature, as well as to show that the way in which **PA** proves these transfinite induction statements involves no other appeal to higher-order notions, then we take that to speak in favor of our conjecture. We will take up these tasks in reverse order: in this section, we shall see that the way in which **PA** deals with transfinite induction claims (in terms of provability) involves no higher-order notions besides the ordinals. Then, in the next section, we shall argue that not even infinite ordinals ordinals are higher-order, as they can be said to follow directly from our understanding of the natural number sequence.

The first question we address is, then: how can it be then that **PA** proves transfinite induction claims, i.e., well-orderings, for *infinite* ordinals? How can we make sense of the fact that the theory of finite mathematics speaks about infinitary objects? It is our belief that the way to approach these questions has to do with the nature of the supremum of all ordinals for which transfinite induction claims are provable in **PA**: ε_0. The point is that the way **PA** deals with sets of size (or lists/sequences/proof trees of length) strictly less than ε_0 does not go beyond the strictly finite, as we shall now see; therefore, they are somehow reachable in a finite way.

In order to clarify what we mean here, we turn to the proof of transfinite induction up to ε_0 in **PA**. Presented for the first time by Gentzen [Gen43], we consider a more up-to-date version by Halbach [Hal14]. The proof in question relies on two lemmas. The first of them is the following:

Lemma 8. $\mathbf{PA} \vdash \mathit{Prog}(\varphi(x)) \rightarrow \mathit{Prog}(\mathcal{J}(\varphi(x)))$

where $\mathrm{Prog}(\varphi(x))$ (that reads '$\varphi(x)$ is progressive') is the universal closure of the formula $\forall \beta < x\, \varphi(\beta) \rightarrow \varphi(x)$, and $\mathcal{J}(\varphi(x))$ is the formula $\forall \xi (\forall \eta < \xi\, \varphi(\eta) \rightarrow \forall \eta < \xi + \omega^x\, \varphi(\eta))$.

And, as for the second lemma:

Lemma 9. *If*

$$\mathbf{PA} \vdash Prog(\varphi(x)) \rightarrow \forall \xi < \alpha\, \varphi(\xi)$$

for all formulas φ of $\mathcal{L}_0$, then

$$\mathbf{PA} \vdash Prog(\varphi(x)) \rightarrow \forall \xi < \omega^{\alpha}\, \varphi(\xi)$$

for all formulas φ of $\mathcal{L}_0$.
NB: these expressions correspond to $\mathtt{TI}(\alpha)$ and $\mathtt{TI}(\omega^{\alpha})$, respectively.

Transfinite induction up to any ordinal below ε_0 can be reached by applying Lemma 2 finitely many times, and Lemma 2 is easily obtainable from Lemma 1. It is thus the latter that requires careful examination. And it is in fact the crux of the proof, for it is where the interweaving with infinite ordinals happens. The formula $\mathcal{J}(\varphi(x))$, sometimes known as Gentzen's jump formula, lies at the heart of this lemma. In all cases in which $x \geq 1$, Gentzen's jump formula seems to announce the possibility of 'infinite jumps'. We can (very informally) understand the jump as stating that, when a given formula φ holds for all ordinals below a given one—finite or not—we can carry that formula along for ω^x-many more numbers above that ordinal. That is, it is as if we were indeed 'jumping' over powers of ω—taking an infinite leap the 'safety' of which (in the sense of well-foundedness) is guaranteed by Gentzen's formula. Notwithstanding these intuitions, we will now argue that these leaps are not infinite after all.

There is, however, a limit to these infinite leaps. This limit is given by Cantor's famous Normal Form Theorem, by which any ordinal below ε_0 can be written as the sum of powers of ω with exponent $< \varepsilon_0$, whereas ε_0 itself and greater ordinals cannot.[10] Since Gentzen's jump formula works exclusively with powers of ω, ε_0 marks the boundary to the number of ordinals we can 'jump over'; hence, even if the jumps were infinite, they could not be of an arbitrarily big number of infinite ordinals. This is also why transfinite induction for ε_0 cannot be proved: the inner structure of Gentzen's formula prevents us from reaching ε_0, and in this we see how pivotal this formula is for the proof. We shall say more about this below.

Now, the other component of Lemma 1 is the notion of 'progressiveness', there abbreviated as *Prog*. To say that a formula is progressive is to say that, when it holds for all ordinals below a given one, it holds for that ordinal. Once we know that a formula is progressive, a transfinite induction claim for some ordinal α is just the assertion that, should the formula be satisfied by 0, progressiveness will carry the formula along the ordinal sequence all the way to α. This is all there is

[10] As is well-known, this becomes very relevant for the choice of an ordinal notation system within a theory of arithmetic.

to transfinite induction, as Gentzen held [Gen69, 291]; therefore, progressiveness is the cornerstone of transfinite induction. Yet the apparent mystery of Lemma 1 in relation to our project is that it shows that Gentzen's jump for a certain formula holds whenever the formula is progressive. That is, the formula is carried along 1 ordinal, and then ω ordinals, and then ω^2 ... and all the way to ω^ω and beyond. As such, the mystery lies in asking how it is possible that a finite, indeed unitary, increment in the satisfaction of a formula along the ordinal sequence can result in increments of the order of powers of ω.

The proof of Lemma 1 gives what we take to be a clear answer to this. If a formula φ is progressive, $\mathcal{J}(\varphi(0))$ holds trivially, for it just expresses that φ is carried one ordinal forward. Informally, **PA** 'sees' the unitary jump as safe (in the sense above, i.e., of well-foundedness)[11]. Now, for $\mathcal{J}(\varphi(x))$ to be progressive, $\mathcal{J}(\varphi(1))$, i.e., $\forall\xi(\forall\eta < \xi\, \varphi(\eta) \to \forall\eta < \xi + \omega\, \varphi(\eta))$, must hold. The key then is that, although we seem to face an ω jump, it is after all a finite one. **PA** is given a certain ordinal ξ as input and has to carry that property for a number of ordinals *below* ω (for whatever η we pick, it will be strictly less than $\xi + \omega$). Hence, **PA** only needs to reiterate what it already 'sees' as a 'safe jump' (the unitary one) a given finite (hence, also safe) number of times. A very similar reasoning goes for $\mathcal{J}(\varphi(2))$: since **PA** 'sees' the ω-jump as safe now, it can perform it once and combine it with a finite number of steps (or perform it twice!) to leap just under ω^2-many ordinals.

In more formal terms, we are performing an outer or external induction on n for ω^n. Likewise, when we consider powers of ω of the form $\omega^\alpha, \omega^\omega > \alpha \geq \omega$, the induction is happening at the next exponential level, i.e. we are performing an induction on n for ω^{ω^n}. The same can be said of any power of ω with exponent $< \varepsilon_0$. Since induction is a perfectly arithmetical task, **PA** can carry out these nested inductions, one after the other, to complete the transfinite induction. Even if the ordinals themselves are infinite, their structure is such that ordinary induction need only be performed a finite amount of times, and so **PA** can deal with it. As mentioned above, the reason why this happens is the fact that these ordinals, when formulated in Cantor's Normal Form, are written down as finite objects. You can see each ordinal below ε_0 as a finite list, the members of which are also finite lists, the members of which are also finite lists... and so on. Therefore, it always remains finite. You can also see the ordinal as a finite tree, the nodes of which are

[11] The reader need not interpret 'sees' here in anything like a model-theoretic sense, as a model that 'thinks' of itself in a certain way (e.g., as containing uncountable objects despite being countable, as given by Skolem's paradox). It is just a very informal way to describe the operations that are going on in **PA** to reach the desired results.

finite trees, the nodes of which are finite trees, etc. In each of these finite objects, we perform regular induction.

As we see it, this explanation seems a step forward in defence of Isaacson's thesis in two different (but related) fronts. First, insofar as it gives an answer to the problem we have raised so far: how can **PA**, as the theory of finite mathematics, prove that certain *infinite* ordinals, i.e., seemingly higher-order objects (a characterisation that we shall later question), are well-founded? Our answer is, then, that we can do it because these ordinals present an inner structure of blended finite strings that **PA** proves well-founded by applying ordinary induction finitely many times. The second front has to do with some remarks presented by Gentzen in his original proof of transfinite induction up to ε_0 in **PA**, for whom the situation was exactly the opposite of the one we have presented here. According to him, for an important segment of the countable ordinals (including ordinals well beyond ε_0), '*transfinite induction* is a form of inference which, *in substance, belongs to elementary number theory*', [Gen69, 307, italics in original] so that '[t]he fact that transfinite induction even up to the number ε_0 is no longer derivable from the remaining number-theoretical forms of inference therefore reveals from a new angle the *incompleteness* of the number-theoretical formalism'(ibid.). In other words, he seems to suggest that transfinite induction for $\alpha \geq \varepsilon_0$ is a genuine, arithmetical statement that is nevertheless independent of **PA**. In Isaacson's terms, it is not 'higher-order'. Recently, Saul Kripke [Kri21] has defended a very similar idea, arguing that $\mathtt{TI}(\varepsilon_0)$ is the first genuine number-theoretic (one can read arithmetical) statement that was shown independent from **PA**. For both Gentzen and Kripke, the unprovability of $\mathtt{TI}(\varepsilon_0)$ is yet another example of the incompleteness of **PA** with respect to arithmetical truth, constituting thus a challenge to Isaacson's thesis. We believe, however, that our account of what underlies transfinite induction for ordinals below ε_0 explains why $\mathtt{TI}(\alpha \geq \varepsilon_0)$, unlike transfinite induction for smaller ordinals, is not properly arithmetical, at least as seen from the way **PA** proves such transfinite induction claims: the inner structure of those ordinals does not allow Gentzen jumps, **PA**'s tool to deal with these claims.

We would like to close this section with a final remark. We have here argued that the treatment **PA** gives to transfinite induction proofs of order less than ε_0 is systematically finitary, and that this is something **PA** cannot carry out with ordinals equal or greater than ε_0. While this dividing line between the finitary and infinitary, to be located well into the infinite ordinals, may initially come as a surprise, it becomes increasingly less so as we learn of different situations where the link between infinite ordinals below ε_0 and finitary mathematics is made explicit. Andreas

Weiermann has investigated some of these examples thoroughly; the following are just two of them:

- The set of ordinals below ε_0, equipped with the usual well-ordering of ordinals, is isomorphic to the set $\mathbb{N}$ with the ordering induced by the so-called Matula numbers (see [Wei05]).

- Weiermann [Wei02] has also shown that the behaviour regarding limit laws (i.e., the probability that any property holds in a structure of arbitrarily large size) for classes of structures of infinite size up to ε_0 is continuous with that for classes of structures of finite size. In particular, when considered as additive systems, these classes of structures meet the so-called zero-one law, that is, all properties have probability either 0 or 1 to be satisfied in structures of arbitrarily large size, whether finite or infinite, as long as the size is less than ε_0. In plain terms: finite structures and infinite structures of size up to, but not including, ε_0, show a certain 'decidability' when it comes to satisfying any property.

3.5 Resisting the nuanced thesis II

In the previous section, we studied how **PA** proves transfinite induction for ordinals below ε_0. The intention was to show that $\mathtt{TI}(\omega^\omega)$ and similar statements does not involve higher-order notions, as far as their **PA**-proof goes, and leaving the ordinals aside. Our focus is now on the ordinals themselves. If we show that they are not higher-order notions, then we pave the way for the conjecture we proposed.

Here, let us briefly digress on the idea of 'higher-order notions'. We previously mentioned that the adjective 'arithmetical' is only applicable, according to Isaacson's work, to true mathematical statements in the language of arithmetic that meet certain conditions. We quickly indicated how to extend that to certain false statements in the language of arithmetic, since we take it that statements such as '5+2=6' count as arithmetical. Moreover, we introduced the term 'higher-order concepts', as including those notions that were part, among others, of proofs of non-arithmetical statements expressible in the language of arithmetic. These include syntactic and infinitary concepts. As a result, we called statements containing these concepts higher-order too. In this fashion, we want to say that a certain class of concepts (not just statements) are 'arithmetical'. After all, we are arguing that not all concepts involved in all proofs are higher-order, so we could say that certain concepts involved in certain proofs are non-higher-order. And, given that the notion of higher-order is defined by opposition to that of arithmetical, it seems that we may just as well label as arithmetical all those non-higher-order concepts. In sum, the idea of a concept being arithmetical is a coherent one in Isaacson's framework seems coherent to us. Hence, we could say that our true enterprise in this section is to argue that ordinals equal or greater than ω and smaller than ε_0 are arithmetical concepts, i.e., are not higher-order.

How can we justify that all ordinals below ε_0 are arithmetical? We believe that, according to the notion of 'arithmetical' that underlies Isaacson's thesis, we just need to show that the concept of the ordinal α, for $\omega \leq \alpha < \varepsilon_0$, follows directly from our understanding of the natural number structure. And we can indeed see that this is the case by considering how we form our conception of the successive infinite ordinals. Begin with ω: it seems to me that we get to make sense of what the ordinal ω ultimately is by considering the successor operation as applied repeatedly to zero, and grasping the limit of that sequence, which we ultimately identify with ω. Once we get an understanding of ω, we can apply the successor operation again, and grasp ω^2. Now, one has learnt how to think of the sequence $\{\omega^1, \omega^2, ...\}$, and can—just like with ω—seize the ordinal ω^ω. Hopefully, this is telling enough to understand how we proceed from ω^ω onward, and all the way up to ε_0.

This account of how we get to form the concept of each ordinal between ω and ε_0 is very much inspired on a passage by Georg Kreisel, where he explains his way of characterising 'visualizable' ordinals—which, after all, is also a discussion on the way we construct, or come to understand, certain ordinals. Thus, writing about finitism and characterising finitist proofs as 'visualizable' proofs, he outlines the kinds of operations we can visualize:

> If one can visualize a structure, then also a sequence of ω copies. (...) [I]f an iteration (of some operation) up to ξ has been visualized, then it can also be visualized to $\xi \cdot \omega$. So much is clear. (...) What one would like to say is this: *if one sees how to visualize iteration up to each $\xi < \xi_0$, then ξ_0 itself can be visualized.* The problem is to put this into formal terms. [Kre65, 170-1]

We take it that, in this passage, Kreisel makes the two remarks on which our account is based: first of all, that if one can visualize a certain structure, obtained through a certain operational procedure, we can visualize ω copies of that structure; secondly, that in visualizing a certain operation being applied n times, and $n + 1$, and $n + 2$, ..., all the way to $n + n$ and more, we can visualize the operation being applied m times, where m is the supremum of that sequence.

Thus, as we suggested, playing with the successor operation and exponentiation in the way Kreisel outlines, we may get to visualize ω^ω and similar ordinals on purely arithmetical grounds (i.e., as following from our understanding of the natural number structure), and so present it as an arithmetical concept. Moreover, Kreisel's complaint that we lack a way to express this phenomenon in formal terms is irrelevant for our purposes. What matters for us is that our understanding of ω^ω was reached employing resources that do not go beyond those already needed in understanding the natural number structure, and so that the given concept is not infinitary in nature for us, but follows from our knowledge of how natural numbers and their operations work. There would be nothing non-arithmetical (in the sense of not following from our grasp of the natural number structure) in $\mathtt{TI}(\omega^\omega)$, nor in $\mathtt{TI}(\alpha < \varepsilon_0)$.

Admittedly, we can cast doubt on Kreisel's believe that this process of visualization is what underlies to *finitist* mathematics (and so did, famously, Tait [Tai81]). But none of that is of our concern either, as long as we can see that this reasoning always remains within the realm of 'finite' (and not 'finitistic') or 'finitary' (as opposed to 'infinitary') mathematics.

A reasonable worry here is that the reasoning could be reiterated all along, and we would end up claiming that ordinals of the order of ε_0 and beyond are of an equally arithmetical nature—perhaps all the way up to the smallest nonrecursive ordinal, ω_1^{CK}. We do not have a clear answer to this. But whatever the stance, we believe that it will not undermine our conjecture. Take one possible way of proceeding here, namely insisting that there is something especial about ε_0 when compared to smaller ordinals. One can point out, for instance, that ε_0 is the minimal fixed-point of the exponential map $\alpha \mapsto \omega^\alpha$. It is also the third exponentially indecomposable ordinal, just after 2 and ω, besides being multiplicatively and additively indecomposable. So when one takes these operations (addition, multiplication and exponentiation) to be the core of the basic operations on natural numbers, ε_0 marks the limit to what we can reach through ω copies attached to a natural number manipulated with one of these operations.

On the other hand, the objector could argue that tetration, pentation and so on are just as elementary a hyperoperation as their lower-level counterparts.[12] And so, the other possible way of proceeding would have it, there is no need to stop at the level of ε_0. One can visualize ordinals all the way up to ω_1^{CK}. In other words, one could construct, on the basis of our understanding of the natural numbers, all ordi-

[12]Actually, the fourth level of the Grzegorczyk hierarchy ($\mathcal{E}^4$) already asserts the totality of tetration; and $\mathrm{I}\Delta_0 + \mathcal{E}^4$ is much weaker than **PA**.

nals the notation of which belongs to what we call Kleene's $\mathcal{O}$ (after [Kle38]). This is a more than sensible move, given the features of the set $\mathcal{O}$. But it does not weaken our conjecture. Remember that we needed to defend that: (1) ordinals below ε_0 are not higher-order concepts, and (2) the non-coded version of the proof of transfinite induction up to ε_0 does not include any other higher-order notions. Insisting that a great deal of ordinals greater than ε_0 are not higher-order either does not undermine our goal, for we explained how proofs of transfinite induction for ordinals equal or greater than ε_0 can no longer rely on non-higher-order concepts alone. In sum: either way of interpreting Kreisel's remarks leads to the same conclusion.

3.6 In search of more evidence

As a reminder, our driving conjecture here is that there is a way to justify the arithmeticality of each statement that may seem higher-order. These include statements about infinitary concepts and those about syntactic concepts.

The argument deployed seems to do well not only with transfinite induction claims but with all statements involving infinitary objects and, in particular, infinite ordinals (for instance, results on ordinal arithmetic; see e.g., [Som95, §3]). Likewise, it seems to us that it fares well with respect to *consistency* statements. But these statements by no means exhaust the class of 'syntactic' statements. For instance, we find that statements that code provability in a theory of arithmetic are of an equally syntactic nature. If we are to defend the conjecture—and, with it, Isaacson's thesis—one will have to tell a convincing story on why these statements are also arithmetical.

For instance, what will we make of Henkin sentences, that is, formulae φ such that

$$\varphi \leftrightarrow \mathrm{Pr}_{\mathrm{PA}}(\ulcorner \varphi \urcorner)?$$

Some considerations come into play here. First of all, there is no one single formula expressing provability in a formal system. The formula in question will depend, among other things, on the choice of coding made, and on the conditions we believe a formula expressing provability in a system should meet. The last point is particularly relevant, and has been the object of some discussion—see e.g. [HV14]. Indeed, if some formula $\pi(x)$ that intends to express provability is generally believed to be unsuccessful for that aim, we are (arguably) no longer talking about a *syntactic* statement, insofar as it fails to capture the relevant syntactic property.

Hence, formulae like the ones Kreisel devised to answer Henkin's problem (i.e., whether Henkin sentences are provable in their relevant systems) [Kre54] might not be strictly relevant when it comes to testing the conjecture: since most would argue that they do not express provability (as Henkin, and Halbach and Visser, have done), they do not contain higher-order notions.

Thus, one could argue that it all boils down to justifying the arithmeticality of Henkin sentences expressed with the 'canonical' formula capturing provability, which we denote as $\mathrm{Bew}(x)$. It is at this point where the defender of the conjecture must step in and try to explain in what sense this type of sentences are arithmetical. We shall not attempt to do that there. Nonetheless, we venture that one can accomplish this task for the formulae in question by identifying provability with the existence of a certain *finite* sequence and, in turn, justifying the arithmeticality of the notion of 'sequence'.

3.7 Final remarks

In this chapter, we introduced the notion of adequacy for arithmetical theories, and showed that there is a reading of Isaacson's thesis under which **PA** can be considered an inadequate theory of arithmetic. As we see it, two possibilities stand out now if such a conclusion is to be avoided. Either we take this to be significant evidence in favour of retaining the Tarskian conception of arithmetical truth as truth in $\mathcal{N}$, going thus back to the incompleteness of **PA**, or we find a way to justify the arithmetical character of statements such as $\mathtt{TI}(\omega^{\omega})$, Con(**PRA**) and the like. Here, we tried to pursue the second path. As we said, our argument is just conjectural, based on a paradigmatic case study, and we do not deny that more may need to be done. But, if the conjecture holds, we believe there is a way to preserve the claim that **PA** is complete, sound *and* adequate with respect to Isaacson's conception of arithmetical truth. As a result, we submit that Isaacson's thesis gains plausibility. And, as we argued, this brings us a step closer to securing the determinacy of arithmetic.

Conclusion

Throughout the thesis, we have repeatedly made clear that this piece of work was always conceived as the initial building bricks of an ongoing, more ambitious project. The driving question for that project, and hence also for this thesis, was: *why is arithmetic determinate, and in what sense—if any—is the arithmetical case different from that of set theory?*

The first chapter of this work intended to clarify and expand on the aforementioned research question. We explained determinacy as a semantic notion related to the possession of a determinate truth-value, and we defined **arithmetical determinacy** as the thesis that all well-formed mathematical statements have a determinate truth-value. We made clear how this work is a response to a challenge that arises from this semantic notion of determinacy, namely the challenge of explaining how the determinacy of arithmetic thus understood arises in the first place. We introduced Putnam's model-theoretic arguments—the argument from Löwenheim-Skolem, the permutation argument, and the constructivisation argument—, and presented them as an important threat to **arithmetical determinacy**. We surveyed influential responses to the Putnamian threats, including the arguments from categoricity (that address the permutation argument), and the arguments from observational constraints (directed against the argument from Löwenheim-Skolem) that Hartry Field popularised. We concluded that these arguments were either somehow unsatisfactory, or under-explored.

Chapter 2 noted that we want to be able to couple the mathematical theories whose determinacy we are exploring with a corresponding theory of determinate truth, in order to raise the questions within the very formal background we work with. We argued that a first step towards finding a theory of determinate truth that can do the job demands understanding supervaluational truth, as well as material adequacy, since we want both of these properties to feature in our theory. We thus assessed and compared three different 'recipes' for supervaluational truth: the duple consisting of van Fraassian semantics and Cantini's **IT**; Stern's semantic **SSK** and the axiomatic **IT**; and McGee's theory of definite truth **MG2**. We offered an in-depth description of the last of these theories, and traced the connections between the minimal fixed points of the three of them—the hope being that this will help us build our own theory of determinate truth in subsequent work.

Finally, chapter 3 resumed the quest for arithmetical determinacy by deploying a way in which the latter can be secured, namely endorsing Isaacson's thesis—the thesis that **PA** is complete and sound with respect to arithmetical truth. In order to defend the thesis, we argued that it can meet an important challenge: the claim that the thesis renders **PA** inadequate as a theory of arithmetic, i.e., that it proves truths that are not arithmetical. Our conjecture, that still needs to be backed further, was that all those truths can be shown to be arithmetical, and we exemplified it by justifying the arithmeticality of transfinite induction and consistency claims. We concluded that this is incipient evidence as to the plausibility of Isaacson's thesis, and so to the determinacy of arithmetic. However, we acknowledged that meeting the metasemantic challenge requires going further, among other things to understand why the arithmetical case is different from the set-theoretic one. It is our ambition to address this and similar far-reaching questions in the near future.

Bibliography

[Bal98] Mark Balaguer. *Platonism and anti-platonism in mathemathics*. Oxford University Press, 1998.

[Bay01] Timothy Bays. On Putnam and his models. *Journal of Philosophy*, 98(7):331–350, 2001.

[BBJ07] George Boolos, John P. Burgess, and Richard Jeffrey. *Computability and Logic*. Cambridge University Press, 2007.

[Bea99] JC Beall. From Full Blooded Platonism to Really Full Blooded Platonism. *Philosophia Mathematica*, 7(3):322–325, 1999.

[Ben65] Paul Benacerraf. What Numbers Could Not Be. *Philosophical Review*, 74(1):47–73, 1965.

[Ben73] Paul Benacerraf. Mathematical truth. *The Journal of Philosophy*, 70(19):661–679, 1973.

[Ber14] Sharon Berry. Malament–Hogarth machines and Tait's axiomatic conception of mathematics. *Erkenntnis*, 79(4):893–907, 2014.

[Ber21] Sharon Berry. Physical Possibility and Determinate Number Theory. *Philosophia Mathematica*, 29(3):299–317, 2021.

[Berng] Sharon Berry. Σ^0_1-Soundness Isn't Enough: Number Theoretic Indeterminacy's Unsavory Physical Commitments. *British Journal for the Philosophy of Science*, forthcoming.

[Bov09] Andrey Bovykin. Exact unprovability results for compound well-quasi-ordered combinatorial classes. *Annals of Pure and Applied Logic*, 157(2-3):77–84, 2009.

[BS12] Tim Button and Peter Smith. The Philosophical Significance of Tennenbaum's Theorem. *Philosophia Mathematica*, 20(1):114–121, 2012.

[Bue05] Otávio Bueno. On the referential indeterminacy of logical and mathematical concepts. *Journal of Philosophical Logic*, 34(1):65–79, 2005.

[Bur86] John P. Burgess. The truth is never simple. *The Journal of Symbolic Logic*, 51(3):663–681, 1986.

[But11] Tim Button. The metamathematics of Putnam's model-theoretic arguments. *Erkenntnis*, 74(3):321–349, 2011.

[BW11] Elizabeth Barnes and J. Robert G. Williams. A Theory Of Metaphysical Indeterminacy. In *Oxford Studies in Metaphysics: volume 6*. Oxford University Press, 2011.

[BW16] Tim Button and Sean Walsh. Structure and Categoricity: Determinacy of Reference and Truth Value in the Philosophy of Mathematics. *Philosophia Mathematica*, 24(3):283–307, 2016.

[BW18] Tim Button and Sean Walsh. *Philosophy and Model Theory*. Oxford University Press, 2018.

[Can90] Andrea Cantini. A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, 55(1):244–259, 1990.

[Ded63] Richard Dedekind. The Nature and Meaning of Numbers. In *Essays on the Theory of Numbers*. Dover, 1963.

[Fef91] Solomon Feferman. Reflecting on incompleteness. *Journal of Symbolic Logic*, 56(1):1–49, 1991.

[Fef08] Solomon Feferman. Axioms for determinateness and truth. *The Review of Symbolic Logic*, 1(2):204–217, 2008.

[Fer16] José Ferreirós. *Mathematical Knowledge and the Interplay of Practices*. Princeton University Press, 2016.

[FFMS00] Solomon Feferman, Harvey M. Friedman, Penelope Maddy, and John R. Steel. Does mathematics need new axioms? *The Bulletin of Symbolic Logic*, 6(4):401–446, 2000.

[FHKS15] Martin Fischer, Volker Halbach, Jönne Kriener, and Johannes Stern. Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, 8(2):257–278, 2015.

[FI19] Salvatore Florio and Luca Incurvati. Metalogic and the Overgeneration Argument. *Mind*, 128(511):761–793, 2019.

[Fie94] Hartry Field. Are our logical and mathematical concepts highly indeterminate? *Midwest Studies in Philosophy*, 19:391–429, 1994.

[Fie98] Hartry Field. Some Thoughts on Radical Indeterminacy. *The Monist*, 81(2):253–273, 1998.

[Fie01] Hartry Field. *Truth and the Absence of Fact*. Oxford University Press, 2001.

[Fie08] Hartry H. Field. *Saving Truth from Paradox*. Oxford University Press, 2008.

[Fin75] Kit Fine. Vagueness, truth and logic. *Synthese*, 30(3/4):265–300, 1975.

[FS87] Harvey Friedman and Michael Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33(1):1–21, 1987.

[G6̈5] Kurt Gödel. *On the length of proofs*, page 82–88. Raven Press, 1965.

[Gas11] Richard Gaskin. Reference and the Permutation Argument. *Proceedings of the Aristotelian Society*, 111:295–309, 2011.

[Gen43] Gerhard Gentzen. Beweisbarkeit und unbeweisbarkeit von anfangsfällen der transfiniten induktion in der reinen zahlentheorie. *Mathematische Annalen*, 119(1):140–161, 1943.

[Gen69] Gerhard Gentzen. *The collected papers of Gerhard Gentzen*. North-Holland, 1969.

[Gö90] Kurt Gödel. What is Cantor's continuum problem? (1964). In Solomon Feferman, editor, *Collected Works Vol. II*, pages 254–270. Oxford University Press, 1990.

[Hal14] Volker Halbach. *Axiomatic theories of truth*. Cambridge University Press, 2014.

[HFng] Volker Halbach and Kentaro Fujimoto. Classical Determinate Truth. *Journal for Symbolic Logic*, forthcoming.

[HH05] Volker Halbach and Leon Horsten. Computational structuralism. *Philosophia Mathematica*, 13(2):174–186, 2005.

[Hor01] Leon Horsten. Platonistic formalism. *Erkenntnis*, 54(2):173–194, 2001.

[Hun02] Edward V. Huntington. A Complete Set of Postulates for the Theory of Absolute Continuous Magnitude. *Transactions of the American Mathematical Society*, 3(2):264–279, 1902.

[HV14] Volker Halbach and Albert Visser. Self-reference in arithmetic i. *The Review of Symbolic Logic*, 7(4):671–691, 2014.

[HW17] Bob Hale and Crispin Wright. *Putnam's Model-Theoretic Argument against Metaphysical Realism*, pages 703–733. John Wiley and Sons, Ltd, 2017.

[Inc08] Luca Incurvati. Too naturalist and not naturalist enough: Reply to horsten. *Erkenntnis*, 69(2):261–274, 2008.

[Inc20] Luca Incurvati. *Conceptions of set and the foundations of mathematics*. Cambridge University Press, 2020.

[Isa87] Daniel Isaacson. *Arithmetical Truth and Hidden Higher Order Concepts*, page 147–169. Elsevier: The Paris Logic Group, 1987.

[Isa92] Daniel Isaacson. *Some Considerations on Arithmetical Truth and the ω-rule*, page 94–138. Routledge, 1992.

[Kle38] Stephen C. Kleene. On notation for ordinal numbers. *Journal of Symbolic Logic*, 3(4):150–155, 1938.

[Kre54] Georg Kreisel. On a problem of Henkin's. *Indagationes mathematicae*, 15:405–406, 1954.

[Kre65] Georg Kreisel. *Mathematical Logic*, volume III, page 95–195. Wiley, 1965.

[Kri75] Saul Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72(19):690–716, 1975.

[Kri21] Saul Kripke. Mathematical Incompleteness Results in First-order Peano Arithmetic: A Revisionist View of the Early History. *History and Philosophy of Logic*, forthcoming, 2021.

[KU08] Philip Kremer and Alasdair Urquhart. Supervaluation fixed-point logics of truth. *Journal of Philosophical Logic*, 37(5):407–440, 2008.

[Kun12] Kenneth Kunen. *The Foundations of Mathematics*. College Publications, 2012.

[Kun13] Kenneth Kunen. *Set theory*. College Publications, 2013.

[Lei05] Hannes Leitgeb. What truth depends on. *Journal of Philosophical Logic*, 34(2):155–192, 2005.

[Lin18] Øystein Linnebo. Platonism in the Philosophy of Mathematics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2018 edition, 2018.

[Mar01] Donald A. Martin. Multiple universes of sets and indeterminate truth values. *Topoi*, 20(1):5–16, 2001.

[McG85] Vann McGee. How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14(4):399–410, 1985.

[McG91] Vann McGee. *Truth, Vagueness, and Paradox: an essay on the logic of truth*. Hackett Pub. Co., 1991.

[McG92] Vann McGee. Maximal consistent sets of instances of Tarski's schema (t). *Journal of Philosophical Logic*, 21(3):235–241, 1992.

[McG97] Vann McGee. How we learn mathematical language. *Philosophical Review*, 106(1):35–68, 1997.

[Mea13a] Toby Meadows. Truth, dependence and supervaluation: Living with the ghost. *Journal of Philosophical Logic*, 42(2):221–240, 2013.

[Mea13b] Toby Meadows. What can a categoricity theorem tell us? *Review of Symbolic Logic*, 6(3):524–544, 2013.

[Mos67] Andrzej Mostowski. Recent results in set theory. In Imre Lakatos, editor, *Problems in the Philosophy of Mathematics: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*, page 82–96. North-Holland, 1967.

[Mos80] Yiannis Nikolaos Moschovakis. *Descriptive set theory*. North-Holland Publ. Co., 1980.

[MR22] JB Manchak and Bryan W. Roberts. Supertasks. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2022 edition, 2022.

[NLeyng] Carlo Nicolai and Mateusz Ł ełyk. A theory of implicit commitment for mathematical theories. *Synthese*, forthcoming.

[NP19] Carlo Nicolai and Mario Piazza. The implicit commitment of arithmetical theories and its semantic core. *Erkenntnis*, 84(4):913–937, 2019.

[Pak22] Fedor Pakhomov. How to escape Tennenbaum's theorem. 2022. https://arxiv.org/abs/2209.00967.

[Par08] Charles Parsons. *Mathematical Thought and its Objects*. Cambridge University Press, 2008.

[Pas14] A. C. Paseau. The overgeneration argument(s): A succinct refutation. *Analysis*, 74(1):40–47, 2014.

[Poh09] Wolfram Pohlers. *Proof Theory. The first step into impredicativity*. Springer, 2009.

[Put80] Hilary Putnam. Models and reality. *Journal of Symbolic Logic*, 45(3):464–482, 1980.

[Put81] Hilary Putnam. *Reason, Truth and History*. Cambridge University Press, 1981.

[Qui70] W. V. Quine. *Philosophy of Logic*. Harvard University Press, 1970.

[Ros21] Lucas Rosenblatt. Should the non-classical logician be embarrassed? *Philosophy and Phenomenological Research*, forthcoming, 2021.

[Say90] Charles Sayward. Four views of arithmetical truth. *The Philosophical Quarterly*, 40(159):155–168, 1990.

[Sch14] Thomas Schindler. Axioms for grounded truth. *The Review of Symbolic Logic*, 7(1):73–83, 2014.

[Sha97] Stewart Shapiro. *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, 1997.

[Sim09] Stephen G. Simpson. *Subsystems of Second Order Arithmetic*. Perspectives in Logic. Cambridge University Press, 2 edition, 2009.

[Smi08] Peter Smith. Ancestral arithmetic and Isaacson's thesis. *Analysis*, 68(1):1–10, 2008.

[Smi13] Peter Smith. *An Introduction to Gödel's Theorems*. Cambridge University Press, 2013.

[Sol06] Robert Solovay. FOM posting, 2006. http://cs.nyu.edu/pipermail/fom/2006-May/010561.html.

[Som95] Richard Sommer. Transfinite induction within peano arithmetic. *Annals of Pure and Applied Logic*, 76(3):231–289, 1995.

[Ste17] Johannes Stern. A note on McGee's ω-inconsistency result. 2017.

[Ste18] Johannes Stern. Supervaluation-style truth without supervaluations. *Journal of Philosophical Logic*, 47(5):817–850, 2018.

[Tai81] William W. Tait. Finitism. *The Journal of Philosophy*, 78(9):524, 1981.

[Tar36] Alfred Tarski. The Concept of Truth in Formalized Languages. In A. Tarski, editor, *Logic, Semantics, Metamathematics*, pages 152–278. Oxford University Press, 1936.

[TB18] Oliver Tatton-Brown. Primitive recursion and Isaacson's thesis. *Thought: A Journal of Philosophy*, 8(1):4–15, 2018.

[Ten59] Stanley Tennenbaum. Non-Archimedean models for Arithmetic. *Notices of the American Mathematical Society*, 6:270, 1959.

[V2̈1] Jouko Väänänen. Tracing internal categoricity. *Theoria*, 87(4):986–1000, 2021.

[vF66] Bas C. van Fraassen. Singular terms, truth-value gaps, and free logic. *The Journal of Philosophy*, 63(17):481–495, 1966.

[vF68] Bas C. van Fraassen. Presupposition, implication, and self-reference. *The Journal of Philosophy*, 65(5):136–152, 1968.

[VW15] Jouko Väänänen and Tong Wang. Internal Categoricity in Arithmetic and Set Theory. *Notre Dame Journal of Formal Logic*, 56(1):121–134, 2015.

[War20] Jared Warren. *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. Oxford University Press, 2020.

[Wei02] Andreas Weiermann. Zero-one law characterizations of ε_0. In Brigitte Chauvin, editor, *Mathematics and computer science II: Algorithms, trees, combinatorics and probabilities*, page 527–539. Birkhäuser, 2002.

[Wei05] Andreas Weiermann. Analytic combinatorics, proof-theoretic ordinals, and phase transitions for independence results. *Annals of Pure and Applied Logic*, 136(1-2):189–218, 2005.

[Wil94] Timothy Williamson. *Vagueness*. Routledge, 1994.

[WV21] Hugh Woodin and Jouko Väänänen. On the categoricity of complete second order theories. *unpublished*, 2021.

[WW20a] Jared Warren and Daniel Waxman. A metasemantic challenge for mathematical determinacy. *Synthese*, 197(2):477–495, 2020.

[WW20b] Jared Warren and Daniel Waxman. Supertasks and arithmetical truth. *Philosophical Studies*, 177(5):1275–1282, 2020.