# *Topics in Arithmetic and Determinacy*

**Nuno Filipe Mendes da Silva Maia**

A thesis submitted for the degree of

*Doctor of Philosophy*




Faculty of Philosophy

University of Oxford

Submitted August 2023

# Abstract

This thesis is about Arithmetical Determinacy. Loosely, this is the problem of whether every question in arithmetic has a determinate answer. In this work I discuss how to exactly understand the concept of determinacy, I criticise arguments for and against the claim that arithmetic is determinate, and examine how questions about determinacy may be applied to other debates in the philosophy of mathematics.

Chapter 1 isolates different ways of understanding the problem of arithmetical determinacy. Chapter 2 turns to mathematical structuralism and explains how popular computability constraints thought to determine the reference of our arithmetical vocabulary are actually unsuccessful in securing determinacy. Chapter 3 criticises an interesting idea for securing determinacy via our experience with supertasks. Chapter 4 explores the phenomenon of mutually inconsistent satisfaction classes and motivates a new account of determinacy in terms of sentences possessing non-classical truth-values. Chapter 5 defends strict finitism, framing some objections against the view in terms of the concept of arithmetical determinacy.

# Acknowledgments

I would like to thank my main supervisor, Joel David Hamkins, for his guidance and for being a model of philosophical and mathematical creativity. I also thank Alex Paseau for his extensive help and advice.

This work is financed by National Funds and by FSE — Fundo Social Europeu — through the Portuguese funding agency FCT — Fundação para a Ciência e a Tecnologia — through grant number SFRH/BD/144393/2019.

# Contents

**Abstract** 1

**Acknowledgments** 2

**Introduction** 6

**1 The Varieties of Arithmetical Determinacy** 9

1.1 Semantic and Ontological Determinacy . . . . . 10

1.2 Zooming in on Semantic Determinacy . . . . . 16

1.2.1 Models, Languages and Theories . . . . . 16

1.2.2 Methodological Constraints . . . . . 24

1.2.3 Intuitiveness and Gödel's First Incompleteness Theorem . . . . . 25

1.3 Structural Determinacy . . . . . 39

1.4 Referential Determinacy . . . . . 42

1.5 Epistemic Determinacy . . . . . 47

1.6 Conclusion . . . . . 48

**2 Redundancy And Non-Classicality: Two Objections To The Argument From Tennenbaum's Theorem** 50

2.1 Introduction . . . . . 51

2.2 Structuralism . . . . . 52

2.3 The Referential Thesis . . . . . 56

2.4 The Argument from Tennenbaum's Theorem . . . . . 58

2.5 Redundancy . . . 60
2.6 Non-Classicality . . . 65
2.7 Conclusion . . . 75

**3 No Semantic Determinacy from Supertask Computations 76**

3.1 Introduction . . . 77
3.2 Semantic Determinacy . . . 78
3.3 From Supertask Computations to Semantic Determinacy . . . 80
3.3.1 Benacerraf and Putnam's Argument, a First Attempt . . . 81
3.3.2 Physical Possibility . . . 84
3.3.3 The '$\omega$th-step' . . . 89
3.3.4 Metaphysical Possibility . . . 91
3.3.5 Taking Stock . . . 94
3.4 Warren and Waxman's Objection . . . 95
3.4.1 The Objection . . . 95
3.4.2 A Refinement . . . 98
3.4.3 Is the $\omega$-rule indispensable? . . . 100
3.4.4 Taking Stock . . . 102
3.5 Model-Relative Computability . . . 102
3.5.1 A New Decision Procedure For Goldbach's Conjecture . . . 103
3.5.2 A New Objection to Benacerraf and Putnam's Argument . . . 105
3.5.3 Replies . . . 108
3.6 Conclusion . . . 116

**4 Krajewski's Theorem, Arithmetical Truth and Paraconsistency 118**

4.1 Introduction . . . 118
4.2 The Problem . . . 120
4.2.1 The No-Difference Thesis . . . 120
4.2.2 No-Difference At Work . . . 123

4.2.3 Satisfaction Classes . . . 126
4.2.4 Krajewski's Theorem and the No-Difference Thesis . . . 130
4.2.5 Taking Stock . . . 135
4.3 A New Reading of Krajewski's Theorem . . . 135
4.3.1 The Idea . . . 135
4.3.2 Interpreting 1/2 . . . 136
4.3.3 Satisfaction and Consequence . . . 138
4.3.4 Subvaluationist Truth . . . 140
4.3.5 Modality and Jaśkowski's Logic . . . 145
4.3.6 Taking Stock . . . 147
4.4 Philosophical Commentary . . . 148
4.4.1 $D_2$, Non-Adjunction and Modality . . . 148
4.4.2 Paraconsistent Arithmetic . . . 151
4.4.3 Revisiting the No-Difference Thesis . . . 152
4.5 Conclusion . . . 154

**5 Is Strict Finitism Arbitrary? 155**
5.1 Introduction . . . 155
5.2 Arbitrariness and Ontological Commitment . . . 156
5.3 Arbitrariness and Facts of the Matter . . . 160
5.4 Arbitrariness and Theory Choice . . . 163
5.4.1 The Charge . . . 163
5.4.2 Knowability: A Reply . . . 165
5.4.3 A Model of Knowability within Strict Finitism . . . 167
5.5 Conclusion . . . 171

**Conclusion 172**

**References 173**

# Introduction

Arithmetic occupies a special role in our theorising about the philosophy of mathematics. A mere cursory glance at the discipline suffices to make clear that twentieth-century scholarship was marked by questions which took the natural numbers as a key case study: debates on the epistemology, ontology and meaning of mathematics were, in several instances, first formulated as debates on the epistemology, ontology and meaning of arithmetic. As the field branches out this trend becomes less and less pronounced. And yet, given the pervasiveness of the natural numbers in reasoning in general, I think it is good methodological policy to let arithmetic be something of a litmus test for general theories in the philosophy of mathematics. In this spirit, in this thesis of five interrelated papers, I propose to study the problem of *Arithmetical Determinacy*. Loosely, this is the problem of whether every question in arithmetic has a determinate answer.

How to precisely formulate the concept of determinacy is already a subject of dispute, and so I should first provide a taxonomy of the different ways of understanding it. This is done in Chapter 1. However, there, I will not aim at neutrality, introducing instead new terminology and advancing possibly controversial distinctions — in this sense, the reader may, if she wishes, regard that chapter as an opinionated introduction to 'the difficult question of how to give a precise and plausible account of the notion of "indeterminacy" ' (Putnam, 2015, 174).

In Chapter 2 I turn to problems about the reference of our arithmetical vocabulary within mathematical structuralism and, in particular, to their connection to Tennenbaum's Theorem. According to one kind of mathematical structuralism, the possible extensions of number-words as well as other arithmetical terms exemplify a unique structure, namely, the structure of the

standard model of Peano Arithmetic. Tennenbaum's Theorem states that any computable model of arithmetic is isomorphic to the standard model. In an attempt to justify the structuralist insight, several authors have argued that interpretations of our arithmetical vocabulary should exemplify computable structures. This assumption together with Tennenbaum's Theorem entails that possible interpretations are isomorphic to the standard model as required. However, I will present two objections to this reasoning. First, I argue that the reasoning trades on an ambiguity in the notion of a computable function which, once spelled out, shows it to be redundant. Second, I argue that the reasoning is not sufficient to prove its conclusion on the basis of there being computable models of non-classical arithmetic which do not exemplify the required structure.

Afterwards, in Chapter 3, I comprehensively criticise an argument suggested by Paul Benacerraf and Hilary Putnam to the effect that our practice with computational procedures describing supertasks suffices to determine the truth of any arithmetical sentence. Their general idea is clear enough. Consider for example Goldbach's Conjecture: the statement that every even number greater than two is the sum of two primes. Assume moreover that we are somehow able to run through all the numbers in a finite amount of time, and mechanically compute whether each number is a counterinstance to the Conjecture by, say, performing each step faster and faster: e.g. it takes half a minute to check the Conjecture for $n = 0$, a quarter to check for $n = 1$, and so on. After a minute, we will have successfully checked all the numbers. Now, if we have not found a counterinstance then there is a fact of the matter – to wit, the result of our computation – securing the sentence's determinate truth; relatedly, if we have found a counterinstance then there is a fact securing the sentence's determinate falsity. Regardless, facts about the result of our computation ensure that the question 'Is Goldbach's Conjecture true?' has a determinate answer. It is in this sense that Benacerraf and Putnam contend that computational procedures describing supertasks suffice to determine the truth of any arithmetical sentence. After strengthening their argument against natural objections, I explain how in many cases computability assertions are themselves model-relative: different models of arithmetic give rise to different conceptions of computation. Building on some of the work from Chapter 2, I

argue that if computational procedures are sufficient for arithmetical determinacy, as Benacerraf and Putnam wish to claim, then the richness in different conceptions of computation yields sentences which are both determinately true and determinately false. I take this to be a refutation of their argument.

Chapter 4 deals with the philosophical significance of satisfaction classes and, in particular, of Krajewski's so-called Non-Uniqueness Theorem. Satisfaction classes provide a formal characterisation of the concept of truth. According to Krajewski's Theorem, there are models of arithmetic admitting mutually inconsistent satisfaction classes. On the surface, this shows that mathematical ontology is not sufficient to pin down the truth of some formulas. This chapter explores a different reading of the Theorem and its consequences. Here, the main idea will be to interpret the phenomenon of mutually inconsistent classes as pointing to the existence of formulas which are both true and false. Based on this idea, a paraconsistent semantics for the notion of 'arithmetical truth' is formulated. It is argued that this reading of Krajewski's Theorem provides a new way of looking at the concept of determinacy where non-classical logics play a key role.

Finally, in Chapter 5, I attempt a partial defence of strict finitism — the view which posits the existence of a largest natural number. Strict finitism is often deemed unconvincing given the arbitrariness over where to locate the largest number along the natural line. Drawing on how arbitrariness arguments are employed in mereology as well as on some distinctions already developed in earlier chapters, I explain how this arbitrariness charge against strict finitism can be understood in several ways. However, I also argue that no matter how the charge is understood, it is always found wanting.

# Chapter 1

# The Varieties of Arithmetical Determinacy

> Curiously, there is not much attention directed at the questions themselves. It is as if everyone already understands what "existence," "object," and "objectivity" come to here. Armed with this understanding, we are ready to take up arms over the matter. What are we fighting about?
>
> Stewart Shapiro, *Philosophy of Mathematics*

**Abstract**

There is no unique way of defining the problem of Arithmetical Determinacy, with different authors intending different theses. Here, I isolate five distinct ways of understanding arithmetical determinacy: what I call semantic, ontological, structural, referential and epistemic determinacy. I argue that most theses are independent of each other.

The problem of Arithmetical Determinacy is, loosely, the problem of whether every question in arithmetic has a 'determinate' answer or, as it is sometimes expressed, the problem of whether every question in arithmetic has an 'objectively correct' or a 'definite' answer.[1] Suppose we ask of some sentence in the language of arithmetic $\varphi$ whether $\varphi$ is true? Call this the

[1] Accordingly, the determinacy of arithmetic is sometimes referred to as the objectivity or definiteness of arith-

$\varphi$-question. A commonplace observation is that we can only answer the $\varphi$-question in a principled way on the assumption that there is some fact of the matter about whether the relevant sentence is true or not. On the contrary, were there no such facts, the proper reaction would be to throw our arms in the air with dejection and complain that the $\varphi$-question does not admit of a determinate answer. This opens the door to a promising way of understanding arithmetical determinacy, in terms of facts of the matter about whether a given sentence in arithmetic is true or not.

## 1.1 Semantic and Ontological Determinacy

Let us say that $\varphi$ is 'determinately true' whenever there is some fact of the matter making it true and no facts of the matter making it false; and that it is 'determinately false' whenever there is a fact of the matter making it false and no facts of the matter making it true; and, finally, that it is 'indeterminate' otherwise.

This account of determinacy is shared by many authors. To give some examples: Putnam (1980, 469) writes that 'the realist standpoint is that there is *a fact of the matter* — a fact independent of our legislation — as to whether $V{=}L$ or not'; for Field (2015, 155) 'the *intended* sense of indeterminacy of truth value is something like: there is no fact of the matter'; and finally, according to Warren and Waxman (2020a, 479), 'we can informally gloss the claim that [the Continuum Hypothesis] is indeterminate as the claim that there is no fact of the matter as to whether or not [the Continuum Hypothesis] is true'. Given this, we can start to isolate one

metic. (For these different uses see Field ([1998] 2001a), Hamkins (2012), and Warren and Waxman (2020a).) Matters are made even more complicated by the fact that some authors prefer to talk about realism instead of determinacy (Putnam, 1980; McGee, 1997; Field, 2015), whilst others use several of these terms within the same discussion with the intent of referring to different concepts (Clarke-Doane, 2020a). Note moreover that my use of 'determinacy' is not to be confused with the very different use that is made of the same word in set theory in connection with the axiom of determinacy.

kind of arithmetical determinacy, what I call 'Semantic Determinacy':

> Every sentence in the language of arithmetic is determinately true or determinately false.
>
> (Semantic Determinacy)

As I noted, I think this is the type of determinacy that many authors have in mind when thinking about whether every arithmetical question has a determinate answer. But how do sentences get to have determinate truth-values? Well, one initial option is that of the realist about mathematical reality who thinks that there exist objects corresponding to the natural numbers.

To explain. On the standard account of satisfiability for arithmetical formulas, satisfaction is defined in terms of a collection of objects — a model's domain — and the way the linguistic expressions occurring in the formula are interpreted in terms of that domain — a model's interpretation function. If the sentence $\varphi$ is interpreted as predicating relations over arithmetical reality — that is, if the sentence $\varphi$ is *about* the existing natural numbers — then, on the assumption that there is a unique way for those objects to be related, $\varphi$ will have a classical truth-value. Under the realist view this is how a fact makes a sentence true[false]: arithmetical reality is[is not] a model for the sentence. As Sayward (1990, 158) writes: 'Classical truth is transmitted ontologically'. In this case, if the objects are related in such a way that $\varphi$ is true then we can answer our starting $\varphi$-question with a 'yes'; if the objects are related in such a way that $\varphi$ is false then we can answer it with a 'no'; either way, the $\varphi$-question admits of a determinate answer. These observations about arithmetical reality are so important for our understanding of semantic determinacy that they deserve a name all of their own, what I call 'Ontological Determinacy':[2]

> There exist arithmetical objects and relations. (Ontological Determinacy)

The argument from the last paragraph may suggest that ontological determinacy entails or is equivalent to semantic determinacy. But this would be confused. To see why, recall that the

[2] Ontological determinacy is what is sometimes called 'realism'. But in view of uniformity, I prefer this new term.

plenitudinous platonist adopts a realist attitude towards every model of first-order Peano Arithmetic. She agrees with the ontological determinacy thesis in that there exist arithmetical objects, but adds that arithmetical reality comes organised in clusters of such objects. Moreover, for each model of arithmetic, there is a cluster such that the relations obtaining among the objects in the cluster correspond to the relations obtaining among the elements in the model's domain. There is then a correspondence between the models and arithmetical reality. Similar to the realist, the plenitudinous platonist thinks that the relevant facts when answering arithmetical questions are those about arithmetical reality. The problem, however, is that in this case there are just too many facts: some of which make a sentence true while others false. This is due to the independence results from Gödel and others which prove the existence of sentences satisfied in some models of arithmetic but falsified in others. Given the correspondence between models of arithmetic and the plenitudinous platonist's arithmetical reality, there will be some clusters of objects which serve as a model for the independent sentence and others that do not — in a way, the sentence is true about some parts of reality and false about others.

How does the plenitudinous platonist go about evaluating ('determining') which sentences are determinate? The natural account is that facts about arithmetical reality make a sentence true whenever *every* cluster is a model for the sentence; and false whenever *every* cluster is not a model for the sentence. Truth is, in a way, the result of the joint contribution of all the clusters. In this case, can we still answer the $\varphi$-question? Well, if every cluster is a model for $\varphi$ then the sentence is determinately true so that we can answer the question with a 'yes'; if every cluster is a model for the sentence's negation then $\varphi$ is determinately false so that we can answer it with a 'no'; otherwise, the question does not have a determinate answer. Finally, note that independent sentences come out as indeterminate. Semantic determinacy is false under plenitudinous platonism, meaning that semantic and ontological determinacy are not equivalent thesis; or, more schematically:

Schema 1

Semantic Determinacy $\not\Leftrightarrow$ Ontological Determinacy

This conclusion is shared by other authors. Indeed, the present section can be thought of as an attempt to develop some of the connections, or lack thereof, discerned in the following passages. First, Field writes that:

> 'Platonism' can mean [...] primarily a thesis about *what exists* [...] But another interpretation of 'platonism' [...] is that higher mathematics is objective in the way that the sciences are: not only is it objective what's a good proof, it's also objective what's a correct axiom, so that e.g. there's an objectively correct answer to the size of the continuum even though this is unsettled by current axioms. But platonism in this sense can go with anti-platonism in the "ontological" sense: witness the Putnam-Hellman idea (Putnam (1967); Hellman (1989)) that mathematical questions are thoroughly objective but to be understood modally. Conversely, ontological platonism needn't require platonism in the "objectivity" sense: witness Mark Balaguer's "plenitudinous platonism" (Balaguer (1998)), or the view that Putnam seems to advocate in the early sections of Putnam (1980), or views on which the existence of mathematical entities and the laws they obey is a matter of convention. (Field, 2016, P-2 (sic))

Also, Berry:

> Anti-Objectivism about set theory can be combined with either realism (i.e., Platonism) or anti-realism about mathematical objects. For example, a philosopher who combines Platonism and anti-Objectivism might say there's a large mathematical universe containing many different hierarchy-of-sets like structures. (Berry, 2022, 190)

In a similar spirit, Shapiro advances a distinction between two kinds of realism — realism about ontology and realism about truth-value — and which comes pretty close to our own distinction between semantic and ontological determinacy:

> There are two different realist themes. The first is that mathematical *objects* exist independently of mathematicians, and their minds, languages, and so on. Call

> this *realism in ontology*. The second theme is that mathematical *statements* have objective truth-values independent of the minds, languages, conventions, and so forth, of mathematicians. Call this *realism in truth-value*. (Shapiro, 1997, 37)[3]

Schema 1 justifies a working assumption implicit throughout the thesis. In the remainder, I will be interested in semantic determinacy (as well as other kinds) almost to the exclusion of ontological determinacy. To this it could be objected that I am blithely ignoring important existence-questions and, therefore, that my discussion of arithmetical determinacy is incomplete or, worse, misguided. However, the preceding discussion already shows that the two topics are to a point independent so that they can be studied separately. I think it is helpful to see Schema 1 'at work'. To close this section I quickly review how ontological and semantic determinacy are relevant to the ongoing debate between the universe and the pluriverse view in set theory.

The problem of determinacy is not local to arithmetic but extends to set theory. We can easily introduce concepts akin to semantic and ontological determinacy but applied to set theory, and rehearse similar examples to yield a schema like Schema 1 (just replace, in the earlier examples, talk of arithmetical reality and arithmetical sentences for set-theoretic reality and set-theoretic sentences). Now, according to the Continuum Hypothesis (CH), every uncountable

[3]Since I will return to these definitions later in the chapter, let me stress that I am wary of *identifying* Shapiro's distinction with our own. This is for two reasons.

First, Shapiro's realism in ontology seems to be closely connected with, if not directly inferred from, extra claims about the reference and meaning of the mathematical vocabulary — e.g. notice the 'thus' in the passage: 'according to model-theoretic semantics, the singular terms of a mathematical language denote objects, and the variables range over a domain-of-discourse. Thus, mathematical objects — numbers, functions, sets, and the like — exist. This is what I call realism in ontology.' (1997, 3–4) Ontological determinacy in my sense is actually independent of those extra claims. For instance, plenitudinous platonism is a form of ontological determinacy which usually interprets a number-word like '2' as a variable ranging over whatever objects play the role of the number 2 in the relevant clusters.

Second, textual evidence suggests that Shapiro's realism in truth-value is compatible with the idea that only the most basic principles in mathematics have objective truth-values — e.g. when describing the Quine-Putnam indispensability argument he writes: 'because mathematics is indispensable for science and because the basic principles of science are (more or less) true, most of the basic principles of mathematics are true as well. This is realism in truth-value'. (1997, 46). Semantic determinacy in my sense states that every sentence in arithmetic is determinate, not only the most basic.

subset of real numbers is bijectible with the whole of reals. So let us ask: is the Continuum Hypothesis the case? Call this the CH-question.

Proponents of the universe view assume a form of realism about sets whereby set theory is understood as describing a *unique* set-theoretic reality of objects – say, the universe $V$ – with set-theoretic statements being about this particular reality. Under realism, since the universe decides the truth-value of CH in one way or another, it is a determinate matter whether the Hypothesis is true or whether it is false. Maddy (2011, 63–64) notes just this: "CH or not-CH' is a theorem [...] as a fact about $V$; therefore CH is either true or false there.' For the universe-theorist, the CH-question admits of a determinate answer: there is a fact of the matter, the universe $V$ and its arrangement, which fixes CH's truth-value.

Still, there might be a wrinkle in the realist's argument. For why think that set-theoretic reality is unique? Why think, that is, that there is only one universe of sets? The multiverse view rejects this assumption, adopting instead a realist attitude towards *every* model of first-order Zermelo-Fraenkel Set Theory (ZF) (see for example Balaguer, 1995, 1998; Hamkins, 2012). Much like the plenitudinous platonist in arithmetic, the multiverse-theorist imagines set-theoretic reality as coming in clusters of sets ('universes'). Moreover, for each model of set theory, there is a cluster such that the relations obtaining among the objects in the cluster correspond to the relations obtaining among the elements in the model's domain. Since CH is independent of ZF, there will be models according to which CH is true and others according to which it is false. Given the correspondence between models and reality, there will be some clusters of objects which serve as a model for CH and others that do not. As a result, the multiverse-theorist goes on to argue, the CH-question does not admit of a determinate answer: there are simply too many equally good models/clusters of set-theoretic reality so that it becomes impossible to fix CH's truth-value in a unique manner.

It is clear that in the universe versus multiverse debate, the former view accepts both the semantic and ontological determinacy theses (suitably understood) whereas the latter accepts the ontological determinacy thesis only. And the relevant point is that had we not been careful enough to distinguish between ontological and semantic determinacy, and had we instead

simply conflated the two theses, this would have trivialised the debate settling it in favour of the universe-theorist. After all, on the assumption that ontological determinacy is equivalent to — and *a fortiori* entails — semantic determinacy, the multiverse view becomes unstable. Hence, the importance of Schema 1 and akin distinctions.

## 1.2 Zooming in on Semantic Determinacy

There is still a lot to say about semantic determinacy. I defined it as the thesis that every sentence in the language of arithmetic is determinately true or determinately false. However, stated thus, it is deceptively simple, raising numerous questions: (1) is there a way to formalise talk of 'facts of the matter'? (2) what do we mean by 'a sentence in the language of arithmetic'? and, for that matter, what do we mean by 'arithmetic'? (3) should good arguments for semantic determinacy take heed of any methodological constraints? and finally, (4) are there any good arguments for the thesis at all? I address questions (1) and (2) in subsection 1.2.1; (3) in 1.2.2; and (4) in 1.2.3.

### 1.2.1 Models, Languages and Theories

#### Intendedness, A First Look

It is common to formalise talk of 'facts of the matter' in terms of 'intended models'. A model of arithmetic is an *intended model* if and only if it captures the relevant facts.[4] For instance, if we follow the realist in holding that the relevant facts for semantic determinacy are those about the

[4] To some readers my terminology might sound idiosyncratic — perhaps too much so. Sometimes talk of an 'intended model' refers instead to the model that (say) 'we want to talk about' or the model that 'we have in mind'; other times 'intended model' and 'standard model' are used interchangeably. Indeed, compare with Balaguer (1998, 60): 'What mathematicians standardly mean when they say that a sentence is true is that it is true in the *standard* model, or the *intended* structure [...] for the given branch of mathematics. [...] The claim that a model (or class of models) is standard is a claim about *us* rather than the model; what is being claimed is that this is the model (or class of models) that is *intended*, that is, that we have in mind with respect to the given theory'. I am not even sure whether this idea of identifying the intended model with 'the model that we want to talk about' is intelligible — as we will see later on, reference is not always as determinate as we would like it to be, especially in the case of mathematics (section 1.4). At any rate this is not what I mean by 'intended model'. My usage takes the prior concept of 'facts of the matter' as constitutive. If my notion of 'intended model' ends up coinciding with that other usage, then that will have been coincidental.

way arithmetical reality is organised then an intended model is one that captures the objects and relations true of that reality: for example, it is one whose domain is just the arithmetical objects and whose relations true in the model are the relations that actually hold among the objects. On the other hand, if we follow the plenitudinous platonist we will think that arithmetical reality comes in clusters and, more importantly, that there is a correspondence between the clusters and the models. To repeat, for each model of arithmetic, there is a cluster such that the relations obtaining among the objects in the cluster correspond to the relations obtaining among the elements in the model's domain. Under plenitudinous platonism, every model becomes intended. With this characterisation in place, arithmetical determinacy can be described in the following way:

For $\varphi$ a sentence in the language of arithmetic:

$$\begin{cases} \varphi \text{ is determinately true if it is satisfied by all intended models} \\ \varphi \text{ is determinately false if it is falsified by all intended models} \\ \varphi \text{ is indeterminate otherwise} \end{cases}$$

If all intended models satisfy $\varphi$, then all the relevant arithmetical facts make $\varphi$ true; therefore, it is determinately true. Similarly, if two intended models disagree on the truth-value of some sentence then the relevant facts do not fix a truth-value for the sentence rendering it indeterminate, as we would expect. Nonetheless, facts are only half the job, given that we still lack a characterization of a sentence in 'the language of arithmetic'. To make progress on this issue, we should first be clear on what options there are when it comes to formulating a language and a theory for arithmetic.

### First and Second-Order Languages

A second-order language will be understood here as an expansion of a first-order language allowing for both relational and functional variables, together with quantification over these. Examples of formulas in a second-order language will include expressions of the form '$X^n(t_1, ..., t_n)$' where the $t_i$s are terms, or '$\exists X^1(\forall x R(x) \rightarrow X^1(x))$' or ... When it comes to the semantics for

second-order languages it is important to distinguish between 'full' (also called 'standard') and 'Henkin' interpretations. The difference, of course, lies in the interpretation assigned to the second-order variables and quantifiers. In first-order logic, the truth of formulas with free-variables is defined relative to a particular variable-assignment. But in the second-order case, the variable-assignment has to provide an interpretation for the new kinds of variables as well. The key point about full semantics is that we define the notion of *variable-assignment* so as to cover 'all possible functions'. That is, in a full model $\mathfrak{M}$ of second-order logic, with a domain $M$, $g$ is a variable-assignment if and only if it assigns to every first-order variable a member of the domain and to every $n$-place relational variable $X$ some subset of $M^n$. We then have the following clauses:

(i) For $X^n$ an $n$-ary relation-variable and $t_1, ..., t_n$ a sequence of terms, $\mathfrak{M}, g \models X^n(t_1, ..., t_n)$ if $\langle t_1^{\mathfrak{M},g}, ..., t_n^{\mathfrak{M},g} \rangle \in g(X^n)$;

(ii) $\mathfrak{M}, g \models \forall X^n \varphi(X^n)$ if $\mathfrak{M}, g' \models \varphi(X^n)$ for every variable-assignment $g'$ agreeing with $g$ at every variable possibly except $X^n$.

On the other hand, Henkin semantics is obtained from similar clauses but by redefining the notion of variable-assignment $g$ to be more restrictive. In particular, in the Henkin case, $g$ assigns to every $n$-place relational variable $X$ some element of $S^n \subseteq \mathcal{P}(M^n)$ — where '$\mathcal{P}(M^n)$' denotes the power set of $M^n$. (Similar distinctions apply to the interpretation of functional variables.) It is readily seen that a full model is just the limiting-case of a Henkin-model where $S^n = \mathcal{P}(M^n)$. Therefore every Henkin-validity is trivially a full-validity. That is, a formula satisfied in every Henkin model is satisfied in every full model, though the reverse does not hold.

Furthermore, every full model is a model of the Comprehension Axiom. That is, the axiom $\exists X^n \forall x_1, ..., \forall x_n (X^n(x_1, ..., x_n) \leftrightarrow \phi(x_1, ..., x_n))$, where $\phi(x_1, ..., x_n)$ does not contain $X^n$ free. Comprehension guarantees that for any formula, there is a relation with the same extension. If, in the axiom's formulation, $\phi$ is allowed to contain bound second-order variables then it is called the Impredicative Comprehension Axiom. Otherwise, if $\phi$ is not allowed to contain bound second-order variables then it is called the Predicative Comprehension Axiom.

Yet, not all Henkin models satisfy Comprehension. If an Henkin model satisfies Impredicative Comprehension then it is called a *faithful* Henkin model. There are recursive deductive systems for second-order logic, containing Impredicative Comprehension, and which extend the first-order logic rules in a natural manner. Faithful Henkin semantic consequence is both sound and complete with respect to these recursive deductive systems (for details see Shapiro, 1991, 65-70 and 89-91).

### First and Second-Order Theories of Arithmetic

With these distinctions at hand, we can look at how theories of arithmetic differ when formalised in first and second-order languages.

**Definition 1.2.1.** *For a first-order language $\mathcal{L}_{PA}$ in the signature $\{0, S, +, \times\}$, Peano Arithmetic (PA) is the first-order theory axiomatised by the following formulas:*

*(i)* $\forall x(S(x) \neq 0)$

*(ii)* $\forall x(x \neq 0 \rightarrow \exists y(S(y) = x))$

*(iii)* $\forall x \forall y(S(x) = S(y) \rightarrow x = y)$

*(iv)* $\forall x(x + 0 = x)$

*(v)* $\forall x \forall y(x + S(y) = S(x + y))$

*(vi)* $\forall x(x \times 0 = 0)$

*(vii)* $\forall x \forall y(x \times S(y) = (x \times y) + x)$

*Together with the axiom-schema for First-Order Induction:*

*(viii)* $(\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(S(x)))) \rightarrow \forall x \phi(x)$

**Definition 1.2.2.** *For a second-order language $\mathcal{L}_{PA_2}$ in the signature $\{0, S\}$, second-order Peano Arithmetic ($PA_2$) is the theory given by axioms (i)–(iii) in 1.2.1 together with the second-order Induction Axiom:*

$$\forall X((X(0) \wedge \forall x(X(x) \rightarrow X(S(x)))) \rightarrow \forall x X(x))$$

PA is a first-order theory[5] and so subject to the usual limitative results of first-order logic, namely, the Compactness Theorem and the Löwenheim-Skolem Theorem, giving rise to pairs

[5] It is worth noting that axiom (ii) in definition 1.2.1 is redundant since it is provable from the remaining axioms plus induction. Yet, its inclusion in our axiomatization of PA allows for a smoother characterization of Robinson's Arithmetic Q as the theory obtained from the Peano axioms minus the induction schema. Since I will have to discuss Robinson's Arithmetic below, I prefer the present axiomatization of PA with the extra axiom (ii).

of non-isomorphic models. On the other hand, full models of $\mathsf{PA}_2$ are pairwise isomorphic. This last result is known as Dedekind's Categoricity Theorem. Since the theorem plays an important role when discussing arithmetical determinacy, it is worth looking at the proof with some detail (see also Shapiro, 1991, 82–83).

**Theorem 1.2.3** (Dedekind). *Let $\mathfrak{M}_1$, $\mathfrak{M}_2$ be two full models with domains $M_1$ and $M_2$. If $\mathfrak{M}_1 \models PA_2$ and $\mathfrak{M}_2 \models PA_2$ then there is an isomorphism $\pi$ between the models $\mathfrak{M}_1 \xrightarrow{\pi} \mathfrak{M}_2$.*

*Proof Idea.* Let $0_1$ and $0_2$ abbreviate the interpretation of 0 under $\mathfrak{M}_1$ and $\mathfrak{M}_2$. Similar abbreviations are introduced for $S_1$ and $S_2$, and for $i_1$ and $i_2$ denoting some arbitrary elements of each model. Define:

$$B = \{A : A \subseteq M_1 \times M_2 \text{ and } \langle 0_1, 0_2\rangle \in A \text{ and } \forall x, y(\langle x, y\rangle \in A \rightarrow \langle S_1(x), S_2(y)\rangle \in A)\}$$

Let $\pi := \bigcap B$. Note that since $M_1 \times M_2 \in B$ and since every other set in $B$ contains $\langle 0_1, 0_2\rangle$, $\pi$ is not empty. It is also clear that $\pi \in B$. We first show that $\pi$ defines a function between the two domains:

$\pi$ is well-defined: We show by induction on $\mathfrak{M}_1$ that for every $i_1 \in M_1$ there is an $i_2 \in M_2$ such that $\langle i_1, i_2\rangle \in \pi$. Define $P := \{i_1 \in M_1 : \exists i_2 \in M_2 : \langle i_1, i_2\rangle \in \pi\}$. For the base case, if $i_1 = 0_1$ then $\langle 0_1, 0_2\rangle \in \pi$ and so $0_1 \in P$. Let $i_1 \in P$. By definition of $P$ there is an $i_2$ such that $\langle i_1, i_2\rangle \in \pi$ and therefore $\langle S_1(i_1), S_2(i_2)\rangle \in \pi$. Hence, $S(i_1) \in P$. By second-order induction in $M_1$, it follows that $M_1 \subseteq P$. And since $P \subseteq M_1$, $M_1 = P$ and $\pi$ is well-defined.

Functionality: We show by induction on $\mathfrak{M}_1$ that if $\langle i_1, i_2\rangle, \langle i_1, j_2\rangle \in \pi$ then $i_2 = j_2$. Define $P := \{i_1 \in M_1 : \exists! i_2 \in M_2 : \langle i_1, i_2\rangle \in \pi\}$. For the base case, suppose $0_1 \notin P$. Then there is a $j_2 \neq 0_2$ such that $\langle 0_1, j_2\rangle \in \pi$. Define $Q = M_1 \times M_2 - \{\langle 0_1, j_2\rangle\}$. We have that $\langle 0_1, 0_2\rangle \in Q$. Also if $\langle i_1, i_2\rangle \in Q$ then $\langle S_1(i_1), S_2(i_2)\rangle \in Q$, since $S_1(i_1) \neq 0_1$ by axiom (i) in definition 1.2.1. It follows that $Q \in B$ but $\langle 0_1, j_2\rangle \notin Q$ contradicting the definition of $\pi$ as the intersection set of $B$. Suppose for induction that $i_1 \in P$; so there

is a unique $i_2$ such that $\langle i_1, i_2\rangle \in \pi$. Now $\langle S_1(i_1), S_2(i_2)\rangle \in \pi$; but suppose also that there is some $j_2 \neq S_2(i_2)$ such that $\langle S_1(i_1), j_2\rangle \in \pi$ (that is, $S_1(i_1) \notin P$). Similar to before, let $Q = M_1 \times M_2 - \{\langle S_1(i_1), j_2\rangle\}$. Then $\langle 0_1, 0_2\rangle \in Q$ since $S_1(i_1) \neq 0_1$. And suppose $\langle m_1, m_2\rangle \in Q$. If $m_1 \neq i_1$, then $\langle S_1(m_1), S_2(m_2)\rangle \in Q$. If $m_1 = i_1$, then, since $i_1 \in P$, it follows that $m_2 = i_2$ and $\langle S_1(m_1), S_2(m_2)\rangle \in Q$ with $S_2(m_2) = S_2(i_2) \neq j_2$. Hence $Q \in B$ but $\langle S_1(i_1), j_2\rangle \notin Q$ which is again a contradiction. By second-order induction in $M_1$, it follows that $M_1 \subseteq P$. And since $P \subseteq M_1$, $M_1 = P$ and $\pi$ is functional.

Arguments similar to before, but using induction in $\mathfrak{M}_2$, also prove that $\pi$ is injective and surjective. Finally, by construction of $\pi$, it is readily seen that $\pi(0_1) = 0_2$ and $\pi(S_1(i_1)) = S_2(\pi(i_1))$. Therefore, $\pi$ is the required isomorphism between the models $\mathfrak{M}_1 \xrightarrow{\pi} \mathfrak{M}_2$. □

The theory of second-order arithmetic is then able to express a unique isomorphism-type, unlike first-order languages where the construction of non-isomorphic models is parasitic. At this point, let us go back to our original question. In the formulation of the semantic determinacy thesis, what do we mean by 'a sentence in the language of arithmetic' and, in fact, by 'arithmetic'? Well, when answering the question we are at a great risk of descending into triviality. To see why, note that a somewhat contentious proposal has it that our conception of natural number is formally captured by full $\mathsf{PA}_2$. (Below I elaborate on its contentious aspect.) It is easy to see that in this case semantic determinacy follows *by fiat*: if 'a sentence in the language of arithmetic' is a sentence in the language $\mathcal{L}_{\mathsf{PA}_2}$ and if, relatedly, 'arithmetic' is just $\mathsf{PA}_2$, then Dedekind's Categoricity Theorem ensures that all models of arithmetic and therefore – note – all *intended* models agree on any sentence's truth-value. That is, any sentence is determinately true or determinately false.

In the remainder, however, I will understand 'arithmetic' to refer to first-order Peano Arithmetic. Of course, the finitist, the ultrafinitist or even the intuitionist may still balk at my assumption. Still, the idea that our conception of natural number includes *at least* the PA-axioms is sufficiently widespread, to the extent that semantic determinacy so understood is an interesting thesis on its own and capable of important theoretical work.

### Reformulating Semantic Determinacy

With all these caveats in place, Semantic Determinacy can be reformulated more conspicuously as:

> Every sentence in $\mathcal{L}_{\mathsf{PA}}$ is determinately true or determinately false. (Semantic Determinacy)

where the 'determinately true/false'-bit is explained in terms of intended models. The claim that arithmetic is semantically determinate comes then very close to the claim that it is (semantically) complete:[6] any sentence is determinately true if it is true in all intended models of the theory, and determinately false if it is false in all intended models of the theory.[7] Obviously, the thesis should be sharply distinguished from the claim that every, or even only some, sentence in $\mathcal{L}_{\mathsf{PA}}$ is determinately true *and* determinately false. The latter is arguably false: if some sentence is determinately true and determinately false then, given our definition of an intended model, all intended models of arithmetic satisfy and falsify that sentence. (Though I concede that if there are no models of arithmetic then all intended models vacuously verify and falsify every sentence.) Though let me note that I will qualify this last statement in Chapter 4; there I will propose that semantic determinacy be defined in terms of the possession of a classical or non-classical truth-value, allowing for arithmetical sentences which are determinately true-and-false.

The preceding yields a semi-formal account of semantic determinacy. By this I mean that while some of the concepts employed — like Peano Arithmetic — are formally regimented, others — like intended models — do not belong to the language of mathematics. I have taken more or less for granted the concept of an intended model. This may be the correct approach

[6] Completeness 'is said in many ways'. Both syntactic and semantic completeness, categoricity and also decidability have all been given the name 'completeness' at some point. The sense of 'completeness' relevant for semantic determinacy is that of semantic completeness, that is: an $\mathcal{L}$-theory $\mathsf{T}$ is said to be semantically complete if and only if for every sentence $\varphi \in \mathcal{L}$, $\mathsf{T} \models \varphi$ or $\mathsf{T} \models \neg\varphi$.

[7] It is curious that again our definitions come close to those of Shapiro's. He writes: 'If $\Phi$ is deductively independent of [a first-order axiomatization of arithmetic] $A$, then there is no "fact of the matter" whether $\Phi$ is true in $A$. Indeed, $\Phi$ holds in some models and fails in others. The sentence $\Phi$ is a legitimate assertion in the theory, and yet it lacks a truth-value. Thus, *realism in truth-value is not delivered without semantic completeness*.' (1997, 140 fn. 27, my emphasis).

since the concept might turn out to be a primitive in our theorizing. Yet, it is not obvious that it is so. Markosian suggests some guiding criteria when deciding if a concept is primitive or, in his words, brutal:

> It seems to me that possessing the following characteristics makes a concept a likely candidate for being assigned the status of brutality in our theorizing: (i) being relatively easy to grasp on an intuitive level, (ii) being such that there seem to be clear-cut cases of both instantiation and non-instantiation, and (iii) being such that no acceptable account of what it is in virtue of which some *xs* instantiate that concept seems to be forthcoming. (Markosian, 1998, 218)

As to (i), the concept of an intended model arises only within a model-theoretic background so the idea that it is graspable at an intuitive level is shaky at best. As to (ii), it is true that models thinking that they themselves are inconsistent represent paradigms of 'unintendedness'; yet, going beyond this simple observation would immediately invite controversy. Finally, (iii) is again not so sure. The literature suggests *tentative* desiderata for a model of arithmetic to be intended, for it to capture the relevant facts (some of which we will encounter throughout the thesis); these include: (a) the model's interpretation of number-words conforms to the meaning singled out by linguistic use; (b) it does not represent incorrect facts about provability; (c) it satisfies a minimality assumption – Carnap's 'extremal axioms' – to the extent that it is an elementary submodel of every other model of arithmetic; (d) it respects certain computability constraints, namely, that the interpretation assigned to arithmetical functions is computable; (e) it correctly represents mathematical reality.

Bottom line: how to best understand intended models, whether they are a primitive, and what criteria the concept should satisfy are all difficult questions, though here I do not have much to contribute by way of elucidation. In the remainder I will have to rely on the reader's informal understanding of this notion.

### 1.2.2 Methodological Constraints

Equipped with a more detailed understanding of semantic determinacy and other kindred concepts, it is time to ask after the methodological constraints which ought to be in place when giving arguments for (and against) the thesis. First, any defence of semantic determinacy should not rely on immoderate assumptions. In this context, the term 'moderation' originates in Putnam (1980) and is characterised by a refusal to postulate 'nonatural mental powers' or 'mysterious perceptions of mathematical objects' (1980, 464–465). (See also Button & Walsh, 2018, chap. 2.) An immoderate argument for semantic determinacy is far too easy: the immoderate can simply appeal to her special faculty of mathematical intuition, her special access to mathematical objects, to justify the existence of a platonic realm of abstract or acausal arithmetical objects and relations; this arithmetical reality will give the relevant facts of the matter responsible for determining the truth-value of any sentence. The difficulty with this argument is obvious. This suppositious faculty of mathematical intuition is a proxy for nothing intelligible, being in tension with a naturalistic account of our cognitive capacities. I think that here Putnam's criticism remains decisive and unforgiving; he writes:

> The Platonist will reply that what this really shows is that we have some mysterious faculty of "grasping concepts" (or "intuiting mathematical objects") and it is *this* that enables us to fix a model as *the* model [...] but this appeal to mysterious faculties seems both unhelpful as epistemology and unpersuasive as science. What neural process, after all, could be described as the perception of a mathematical object? Why of *one* mathematical object rather than another? (Putnam, 1980, 471)

This is not to say that satisfactory arguments for semantic determinacy must refrain from positing abstract acausal objects. But it is to say that those objects should not play an explanatory role which contradicts a broadly naturalistic outlook of the human mind. (For a similar remark see Warren & Waxman, 2020a, 483–484.)

Moreover, it is believed that our cognitive capacities involve only the computation of effective functions from state inputs to outputs (see Van Rooij (2008) and the references therein)

and, therefore, are bounded by the existence of effective procedures. By the Church-Turing Thesis — every effectively computable function is Turing computable — this means that human capacities form a subset of the Turing computable functions. As McGee (1997, 117) writes, 'human beings are [...] finite systems whose behavioral responses to environmental stimuli are produced by the mechanical operation of natural forces. Thus, according to Church's Thesis, human behavior ought to be simulatable by a Turing machine'. In fact, perhaps, our limitations are even more pronounced for it is sometimes assumed that human capacities are bounded by functions all included in a particular complexity class, the leading intuition here being that actual physical constraints on the human brain (for instance, the number of neurons being at most $10^{12}$) make it the case that human capacities cannot involve functions running in super-polynomial time. Some authors then conjecture that the bound is given by functions running in polynomial time (Frixione, 2001) while others, more liberally, propose that human capacities are limited by functions which are fixed-parameter tractable (Van Rooij, 2008). (Moreover, a function $f$ is of polynomial time if it runs in time $O(n^{\alpha})$ for constant $\alpha$, where big-$O$ defines an asymptotic upper bound; and a function $f$ is fixed-parameter tractable if it runs in $O(g(k)\cdot n^{\alpha})$ for parameter $k$ and constant $\alpha$.) If correct, they justify a second methodological constraint which proscribes the attribution to humans of cognitive powers not replicable by Turing machines.

### 1.2.3 Intuitiveness and Gödel's First Incompleteness Theorem

#### Intuitiveness

So far so good. But are there actually any good arguments for semantic determinacy? Well, the first thing to note is that the thesis is often deemed markedly intuitive, being the default position in these debates. Let me give some testimonials. In his classic 'Which undecidable mathematical sentences have determinate truth values?' Field observes:

> I think that anti-objectivism has considerable plausibility for the typical undecidable sentences of set theory. It has much less plausibility for the undecidable sentences of elementary number theory: these strike almost everyone pretheoretically

> as having determinate truth value, though we may not know what it is. (Field, [1998] 2001b, 337)

Also discussing set theory, Balaguer thinks that:

> Mathematicians are almost universally convinced that objectivism is true in arithmetic, i.e., that all arithmetical sentences are either correct or incorrect. But many of these same mathematicians – indeed, I think we can say most – harbor doubts about whether this is the case in set theory. (Balaguer, 2001, 106)

And when comparing the models of geometry to those of arithmetic, Gaifman notes:

> As far as *Euclidean* geometry is concerned, [the models of non-Euclidean geometry] are not models at all, since they violate one of the axioms. But they have been recognized as models of a different kind of geometry, with a legitimate claim on the concept of a geometric space. [...] By contrast, a non-standard model of arithmetic has no analogous claim on the concept of natural number; it is not as if we have become aware that 'the sequence of natural numbers' has another legitimate interpretation – one which falsifies a standard arithmetical truth. (Gaifman, 2003, 12)

Referring to the structure of arithmetic, Feferman argues:

> The conception of $(\mathbb{N}^+, 1, Sc, <, +, \times)$ is so clear that, again implicitly, at least, there is no question in the minds of (amateur or professional) number theorists – nor should there be any question in our minds – as to the definite meaning of such statements and the assertion that they are true or false, independently of whether we can establish them one way or the other. In other words, in more modern terms, we should accept realism in truth values for statements about this structure [...]' (Feferman, 2011, 14–15)

More recently, Warren writes:

> The determinacy of standard mathematical claims, especially the claims of basic arithmetic, is intuitively plausible. It is hard to even imagine what it would be like for arithmetic to be indeterminate—how could it be indeterminate whether there is a largest twin prime pair? What kind of borderline case could there possibly be? The difficulty of understanding how arithmetical claims could possibly be vague or indeterminate is one reason for thinking they aren't. (Warren, 2020b, 204–205)

And, finally, Picollo and Waxman:

> When we use the language of arithmetic, it is natural to suppose that we have a unique structure in mind: *the natural number structure*. As a consequence, it is natural to regard any arithmetical question as yielding a definite answer. Are there infinitely many prime numbers that differ by 2? The twin prime conjecture famously answers 'yes'. Although we have not currently proved or refuted the twin prime conjecture, it is intuitively hard to see how it could lack a determinate truth-value. After all, if we could run through the natural numbers, we would ultimately find out the answer. Of course, there are infinitely many numbers, so we cannot actually run through them; but this looks like a medical impossibility, not a mathematical one. (Picollo & Waxman, in press, 2)[8]

Nonetheless, these testimonials should be taken with a grain of salt. As I see it, the reason for finding semantic determinacy an intuitive thesis is due to how arithmetic contrasts with other mathematical domains. An example will help. I noted earlier that ideas similar to the semantic determinacy thesis are also discussed in set theory and, in particular, in connection to the Continuum Hypothesis: is CH the case? does CH have a determinate truth-value? We might agree that it is either true or false relative to a model of set theory but – and this is the important 'but' – several authors would demur at a similar claim which dispensed with this sort of relativization, arguing instead that the idea that CH is indeterminate strikes them as more

[8] This idea of being able to determine the truth-value of some arithmetical sentence by 'running through the natural numbers' will be the subject of Chapter 3.

intuitive. By way of explanation they would perhaps point to the richness of our experience with different set-theoretic models. Indeed, Hamkins (2012, 2015) argues that the independence results from Gödel and Cohen, the method of forcing, and other technical developments have by now provided set-theorists with a deep understanding of how to obtain CH or its negation in different models of set theory. This, he goes on to argue, grounds set-theorists' intuitions on the indeterminacy of CH, leading them to regard as unnatural an axiom that decides CH in any particular way.[9] Now, because an analogue to CH is not so easily found in the arithmetical case, semantic determinacy in arithmetic may look intuitive. But for Hamkins this impression is contingent on present technical developments:

> My long-term expectation is that technical developments will eventually arise that provide a forcing analogue for arithmetic, allowing us to modify diverse models of arithmetic in a fundamental and flexible way, just as we now modify models of set theory by forcing, and this development will challenge our confidence in the uniqueness of the natural number structure, just as set-theoretic forcing has challenged our confidence in a unique absolute set-theoretic universe. (Hamkins, 2012, 428)

I think this should give us pause when employing intuitions in arguments *for* semantic determinacy. Besides, we actually find in the philosophy of mathematics arguments *against* it.

### The Argument From Gödel's First Incompleteness Theorem

A somewhat folkloric argument has it that Gödel's First Incompleteness Theorem actually produces an indeterminate sentence. For a consistent theory, the First Incompleteness Theorem lays down a set of sufficient conditions for producing an independent sentence. As long as the theory is also computably axiomatizable, sufficiently expressive – in the sense of being able to define every primitive recursive predicate – and $\omega$-consistent there is a mechanical procedure for

[9] To be sure, I am reading Hamkins as advancing an explanation as to why it is intuitive that CH is indeterminate, and not as advancing an argument for the claim that CH is indeterminate *tout court*. Though Hamkins seems to take set-theorists' experience as an argument for the latter.

constructing a Goldbach-type sentence such that neither it nor its negation is provable from the theory's axioms alone. Perhaps more surprising than the existence of independent sentences altogether, is that computably axiomatizable and sufficiently expressive theories already appear for very weak fragments of arithmetic. Robinson's Q (the theory obtained from the PA-axioms minus the Induction Schema) is expressive enough to define any computable predicate, including functions built via minimization. It follows that PA, on the extra assumption that it is also $\omega$-consistent, satisfies the conditions for Gödel's First Theorem. (Of course, the strengthened version due to Rosser (1936) does without the $\omega$-consistency assumption; for an up-to-date survey of current research on Gödel's Theorems see Cheng (2021).)

Consider some independent sentence. For vividness, suppose we have fixed a particular coding system and an independent sentence of the form $\mathsf{Gödel}_{\mathsf{PA}} := \neg\exists \mathsf{x}\mathsf{Proof}(\mathsf{x}, \overline{\ulcorner \mathsf{Gödel}_{\mathsf{PA}} \urcorner})$ such that $\mathsf{PA} \nvdash \mathsf{Gödel}_{\mathsf{PA}}$ and $\mathsf{PA} \nvdash \neg\mathsf{Gödel}_{\mathsf{PA}}$. The argument against semantic determinacy now claims that facts about arithmetical proof are silent on the truth of the Gödel-sentence and, more generally, that there are two intended models of first-order arithmetic $\mathfrak{M}_1$ and $\mathfrak{M}_2$ such that the former satisfies $\mathsf{Gödel}_{\mathsf{PA}}$ while the second its negation. And because there are intended models disagreeing on the truth-value of $\mathsf{Gödel}_{\mathsf{PA}}$ it follows that the sentence is neither determinately true nor determinately false. $\mathsf{Gödel}_{\mathsf{PA}}$ is indeterminate. Now, this is perhaps not the best of arguments but it is important to understand why we might think so. Thus, I move to survey some possible replies.[10]

[10] I should note that the Gödel-sentence for PA is being used here merely as illustrative of how a case against determinacy might be built. Indeed, one possible reply to this argument for indeterminacy states that the proof of the First Incompleteness Theorem is also a proof that $\mathsf{Gödel}_{\mathsf{PA}}$ is true (in the standard model) and so, presumably, determinate. I am happy to concede the point. Indeed, if the reader finds this reply compelling, it suffices for her to replace reference to the Gödel-sentece in this subsection with another independent sentence whose truth might not be so certain. For example, the Paris-Harrington theorem gives us yet another independent sentence. However, recognising the truth of this new sentence involves conceptual resources — like recognising that Köning's Lemma is true — which seem to exceed the conceptual resources involved in the grasp of elementary arithmetical truths (for elaboration on this point see Isaacson ([1987] 1996), Smith (2013, 225) and the references therein).

### Reply 1: Dedekind's Categoricity Theorem

Shapiro (1991) argues for a formalisation of our concept of natural number in terms of full $\mathsf{PA}_2$. This idea is embedded in a more general hypothesis according to which mathematicians employ concepts definable only in a higher-order setting (e.g. a categorical characterization of the ordered field of real numbers in terms of Dedekind's Completeness Axiom, or cardinality concepts like a Dedekind-Infinite set) and implicitly assume the full range of their second-order variables, so that the adoption of full second-order arithmetic is no more problematic than classical informal mathematics: '[...] the assumptions of second-order logic may be more conspicuous than those behind the practice of classical mathematics, but they are not more troublesome' (1991, 218). Implicit in the argument is a mild form of naturalism whereby in some cases mathematical practice takes precedence over philosophical theorising. In the present case, mathematical practice informs the correct formalisation of our concept of natural number.[11] If correct, Shapiro's view easily lends itself to an argument for semantic determinacy: if the formalisation of the natural number concept takes place in a higher-order theory like $\mathsf{PA}_2$, Dedekind's Categoricity Theorem will secure the determinacy of any sentence.[12]

Yet, this cannot be the whole story. After all, Dedekind's Theorem is sensitive to the semantics adopted: it holds only when we restrict our attention to full models. With full models the universal quantifier in the second-order Induction Axiom is interpreted as ranging over the full power set of the model's domain. The problem in this case is that the set-theoretic metatheory laying down a formalization of the power set concept admits, if first-order, models of arbitrary infinite cardinalities. It follows that the interpretation given to the 'power set of the natural numbers' will change drastically according to the model of set theory available in the

[11] This (somewhat vague) form of naturalism is hinted at in later work: 'Philosophy of mathematics does not supply first principles for mathematics, nor does philosophy operate from a privileged, more secure perspective. Mathematics and philosophy of mathematics are interlocked but autonomous disciplines. In philosophy, we try to interpret the language of mathematics and try to understand its role in the intellectual enterprise' (Shapiro, 1997, 51).

[12] I should stress that here I am assuming a set-based semantics for second-order logic. However, there are other approaches which do not resort to set theory like a plural semantics. Since plural logic will not play a role in the chapters that follow, I will ignore this option — though for discussion see Florio and Linnebo (2016, 2021).

background — e.g. if the model for the metatheory is countable then the 'power set of the natural numbers' will be interpreted as a countable set; in some other uncountable models it will be interpreted as an uncountable set. And the philosophical upshot is that we should be wary of founding the determinacy of arithmetical truth on less secure concepts like power set, especially when the determinacy of CH is already so disputable. (To my knowledge, this argument ultimately traces back to Weston (1976, especially 288–289).) Compare with: 'Although perhaps in the end we should not be bothered by the use of less elementary means to prove the uniqueness of the natural numbers, we cannot just accept them without examination' (Parsons, 1990, 14); and with 'it begs all the questions at issue to suppose that second-order assertions are all semantically determinate, just as it would beg all the questions at issue to suppose that first-order set-theoretic sentences are all semantically determinate' (Field, 2015, 149–150).

For a related point, it is easy to see that Dedekind's Theorem does not hold in Henkin semantics. Here the range of the second-order quantifiers is restricted to a particular subset of the model's domain. With Henkin semantics there is no guarantee that induction can be applied to important properties like $\{i_1 \in \mathcal{M}_1 : \exists i_2 \in \mathcal{M}_2$ such that $\langle i_1, i_2\rangle \in \pi\}$, essential in proving that $\pi$ defines an isomorphism between two $\mathsf{PA}_2$-models $\mathfrak{M}_1$ and $\mathfrak{M}_2$. Furthermore, faithful Henkin semantic consequence is Lindström-equivalent to first-order logic in the sense that a Compactness and Löwenheim-Skolem Theorems can also be proved for it. In this case, non-isomorphic pairs of models are easily reintroduced barring categoricity. This means that even after circumventing the issues with the interpretation of 'power set of the natural numbers', the advocate of Dedekind's Theorem has still to explain why second-order quantification should not be interpreted Henkin-style. As Putnam (1980, 481) writes in a well-known passage: 'the "intended" interpretation of the second-order formalism is not fixed by the use of the formalism (the formalism itself admits so-called "Henkin models", i.e., models in which the second-order variables fail to range over the full power set of the universe of individuals), and it becomes necessary to attribute to the mind special powers of "grasping second-order notions"'.

Finally, we noted that there is a *recursive* deductive system sound and complete with respect to faithful Henkin semantic consequence. By Gödel's Theorem, no such recursive system

exists for full semantic consequence. Full consequence is computationally rich, making us wonder how we can understand it without violating the methodological constraints about human minds supposed to be bounded by Turing machines. (See for example Picollo and Waxman (in press, 4–6) for a recent and clear statement of this issue.)

### Reply 2: Internal Categoricity

The sensitivity of Dedekind's Categoricity Theorem to the interpretation of the higher-order variables pours cold water on the plan to secure semantic determinacy via second-order arithmetic. On the surface, this difficulty would be evaded by a categoricity result which avoided reference to higher-order vocabulary and which could be proved for a theory formulated in a first-order language. And as it turns out, second-order Induction is dispensable for categoricity as long as we allow for a strengthening of the first-order Induction Schema with formulas in an expanded language enabling reference to the elements of different number-structures. This was originally observed by Parsons:

> [B]y looking a little more closely at Dedekind's theorem and its proof. The main point is that it is essentially *first-order*. [...] What is essential to the proof that the two structures are isomorphic is not any rich system of subsets of the one or the other, but rather that it is possible to apply induction and recursion within one of them in a way that involves reference to the other. The usual language of first-order arithmetic only talks about the natural numbers; the functors and predicates operate within the natural numbers. Thus it does not contain the predicates induction on which would show that some independently described structure is isomorphic to it. (Parsons, 1990, 34–35)

To elaborate on this proposal, define:

**Definition 1.2.4.** *For a language of first-order arithmetic* $\mathcal{L}_n = \{0_n, S_n, +_n, \times_n\}$ *define* $\mathsf{PA}(\mathcal{L}_n)$ *to be the theory comprised of the usual first-order Peano Axioms together with the infinitely-many instances of the first-order Induction Schema, as in Definition 1.2.1. We let* $\mathsf{PA}(\mathcal{L}_n) + \mathsf{Ind}_{\mathcal{L}_n \cup \mathcal{L}_m}$ *be just*

*like* $\mathsf{PA}(\mathcal{L}_n)$ *but where now instances of the Induction Schema include formulas in the expanded language* $\mathcal{L}_n \cup \mathcal{L}_m$.

A model for $\mathsf{PA}(\mathcal{L}_n) + \mathsf{Ind}_{\mathcal{L}_n \cup \mathcal{L}_m}$ is just like a model of PA but which (possibly) satisfies *more* instances of Induction than the instances definable in $\mathcal{L}_n$.

**Theorem 1.2.5** (Internal Categoricity. Parsons (1990); Väänänen (2021))**.** *Consider two languages of arithmetic* $\mathcal{L}_1 = \{0_1, S_1, +_1, \times_1\}$ *and* $\mathcal{L}_2 = \{0_2, S_2, +_2, \times_2\}$. *Then, where* $D_1$ *refers to the deductive system of first-order logic, it follows:*

$$PA(\mathcal{L}_1) + Ind_{\mathcal{L}_1 \cup \mathcal{L}_2} \cup PA(\mathcal{L}_2) + Ind_{\mathcal{L}_1 \cup \mathcal{L}_2} \vdash_{D_1} ISOM(0_1, 0_2, S_1, S_2, +_1, +_2, \times_1, \times_2)$$

*where ISOM abbreviates a first-order statement defining an isomorphism which preserves the constant and function-symbols in* $\mathcal{L}_1 \cup \mathcal{L}_2$.

There is a sense in which Theorem 1.2.5 is not about models of arithmetic nor does it presuppose any other semantic notions. On the contrary, it establishes the existence of a purely syntactic object, namely, a deduction in first-order logic. That is why it is called an *internal* categoricity result. Internal categoricity seems to open up a route to semantic determinacy circumventing the earlier difficulties with the interpretation of the second-order variables. For does it not already show that all models of PA are isomorphic without having to quibble over second-order quantification? A moment's thought reveals otherwise. The theory $\mathsf{PA}(\mathcal{L}_1) + \mathsf{Ind}_{\mathcal{L}_1 \cup \mathcal{L}_2} \cup \mathsf{PA}(\mathcal{L}_2) + \mathsf{Ind}_{\mathcal{L}_1 \cup \mathcal{L}_2}$ is a first-order theory and, therefore, it is still subject to first-order logic's limitative results. Assuming it has an infinite model, it also has models of arbitrary cardinalities.

This points to an affinity between Dedekind's Theorem and Theorem 1.2.5, stressing even more the limitations of internal categoricity in securing semantic determinacy. Put aside the issue with Henkin semantics, and assume that 'all' is given a full interpretation. Does Dedekind's Theorem now show that arithmetic is determinate? Not quite. The set-theoretic metatheory where Dedekind's Theorem is proved still admits, if first-order, non-isomorphic models with possibly different sets of 'natural numbers'. That is, for two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ of the metatheory, and two sets $\langle \omega_1, \mathrm{S}, 0 \rangle$ and $\langle \omega_2, \mathrm{S}, 0 \rangle$ standing for the natural numbers in those models

respectively, Dedekind's Theorem ensures that any model in $\mathfrak{M}_1$ satisfying second-order arithmetic is isomorphic to $\langle \omega_1, \mathrm{S}, 0\rangle$ — a similar claim for the sets in $\mathfrak{M}_2$. However, nothing follows from this as to $\langle \omega_1, \mathrm{S}, 0\rangle$ and $\langle \omega_2, \mathrm{S}, 0\rangle$ being isomorphic. We still have possibly non-isomorphic candidates for the natural numbers. In this case all that the categoricity result is able to show is that *within* each model of the metatheory arithmetic is categorical. And we can read Theorem 1.2.5 in a similar way, though this time the problem with different natural number candidates already arises for the models of the object-theory. Let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ now be two models of $\mathsf{PA}(\mathcal{L}_1) + \mathsf{Ind}_{\mathcal{L}_1\cup\mathcal{L}_2} \cup \mathsf{PA}(\mathcal{L}_2) + \mathsf{Ind}_{\mathcal{L}_1\cup\mathcal{L}_2}$, and $\mathfrak{M}_1 \restriction \mathcal{L}_1$ and $\mathfrak{M}_2 \restriction \mathcal{L}_1$ the respective reducts. The latter can be thought as standing, in the language of $\mathcal{L}_1$, for the natural numbers existing inside $\mathfrak{M}_1$ and $\mathfrak{M}_2$ respectively. Moreover, we know that the $\mathcal{L}_2$-reduct of $\mathfrak{M}_1$ is isomorphic to $\mathfrak{M}_1 \restriction \mathcal{L}_1$ from the perspective of $\mathfrak{M}_1$ — a similar claim for the $\mathcal{L}_2$-reduct of $\mathfrak{M}_2$. However, nothing follows from this as to $\mathfrak{M}_1 \restriction \mathcal{L}_1$ and $\mathfrak{M}_2 \restriction \mathcal{L}_1$ being isomorphic. Nor, for that matter, do $\mathfrak{M}_1$ and $\mathfrak{M}_2$ need to be isomorphic. We conclude that first-order internal categoricity does not entail the semantic determinacy thesis.[13]

But perhaps these limitations are due not to *internal* categoricity but to *first-order* internal categoricity. Indeed, second-order arithmetic also enjoys an internal categoricity result of its own. To explain. The existence of a number-structure is normally expressed in terms of a semantic clause, that is, in terms of a set $\mathfrak{M}$ such that $\mathfrak{M} \models \mathsf{PA}_2$. But unlike PA, the theory $\mathsf{PA}_2$ is finitely axiomatizable. We can then consider a second-order sentence formed from the finite conjunction of the $\mathsf{PA}_2$-axioms relativised to a predicate-constant, in the following way:

**Definition 1.2.6.** $PA_2^{int}(N, S, 0) =$

$$N0 \ \wedge \ \forall x(Nx \rightarrow NS(x)) \ \wedge$$

$$\forall x(Nx \rightarrow \neg(S(x) = 0)) \ \wedge \ \forall x\forall y((Nx \wedge Ny) \rightarrow (S(x) = S(y) \rightarrow x = y)) \ \wedge$$

$$\forall X((X0 \wedge \forall x(Nx \rightarrow (Xx \rightarrow XS(x)))) \rightarrow \forall x(Nx \rightarrow Xx))$$

According to $\mathsf{PA}_2^{\mathrm{int}}(N, S, 0)$, the predicate $N$ behaves very much like a number-structure

[13] I hasten to add that the arguments in the main text are not meant as a criticism of Parsons's (1990) paper. Indeed, there, Parsons appeals to internal categoricity in order to justify the claim that, loosely, two agents agreeing on the Peano axioms will agree on which formulas follow from those axioms. This is different from the semantic determinacy thesis and, therefore, my arguments do not extend to this other project by Parsons.

obeying the standard principles for second-order arithmetic. Button and Walsh (2018, chapters 10 and 12) go on to propose that number-structure talk as well as other semantic notions, including categoricity and determinacy, be understood not in terms of models but in terms of formulas and proofs in the language of second-order logic:

> **The internalist manifesto.** For philosophical purposes, the metamathematics of second-order theories should not involve semantic ascent. Instead, it should be undertaken within the logical framework of [the] very theories under investigation. Our slogan is: METAMATHEMATICS WITHOUT SEMANTICS! (Button & Walsh, 2018, 227)

Following their terminology, call a sentence like $\mathsf{PA}_2^{\text{int}}(N, S, 0)$ an 'internal-structure'. The existence of an isomorphism between internal-structures can also be expressed in second-order language. That $F$ is an isomorphism between two internal-structures is the conjunction:

**Definition 1.2.7.** $ISOM(F, N_1, N_2, S_1, S_2, 0_1, 0_2) :=$

$$\forall x, \forall y(F(x, y) \to (N_1(x) \wedge N_2(y))) \quad \wedge \quad \forall x(N_1(x) \to \exists! y F(x, y)) \quad \wedge$$
$$\forall y(N_2(y) \to \exists! x F(x, y)) \quad \wedge \quad F(0_1, 0_2) \quad \wedge \quad \forall x \forall y(F(x, y) \to F(S_1(x), S_2(y)))$$

The standard recursive deductive system $D_2$ for second-order logic, which includes the Impredicative Comprehension Axiom, proves an internal categoricity result in the following sense:[14]

**Theorem 1.2.8** (Internal Categoricity. Väänänen and Wang (2015))**.**

$$\vdash_{D_2} \forall N_1, N_2, S_1, S_2, 0_1, 0_2((PA_2^{int}(N_1, S_1, 0_1) \quad \wedge \quad PA_2^{int}(N_2, S_2, 0_2)) \to \exists F \ \ ISOM(F, N_1, N_2, S_1, S_2, 0_1, 0_2))$$

If structure talk is understood in terms of internal-structures, Theorem 1.2.8 tells us that number-structures are pairwise isomorphic. Even though this might remind us of Dedekind's Theorem, unlike it, internal categoricity for second-order arithmetic is a purely syntactic result:

[14] Impredicative Comprehension is essential in this case. Theorem 1.2.8 does not follow with the predicative version only, as proved in Button and Walsh (2018, 246ff).

it simply states that there is a particular deduction in $D_2$. (Still, it is interesting to note that the result can be read semantically: since faithful Henkin semantic consequence is sound and complete with respect to $D_2$, the Theorem also tells us that models of $\mathsf{PA}_2$ within a common faithful Henkin structure are witnessed to be isomorphic by that larger structure.)

For our purposes, what is important is not so much Theorem 1.2.8 and instead the following corollary stating, in essence, that internal-structures do not tolerate indecision:

**Corollary 1.2.8.1** (Intolerance. Button and Walsh (2018)). *For any formula $\varphi(N_1, S_1, 0_1)$ whose quantifiers are $N_1$-restricted and whose free variables are all displayed:*

$$\vdash_{D_2} \forall N_1, S_1, 0_1(PA_2^{int}(N_1, S_1, 0_1) \rightarrow \varphi(N_1, S_1, 0_1)) \vee \forall N_1, S_1, 0_1(PA_2^{int}(N_1, S_1, 0_1) \rightarrow \neg\varphi(N_1, S_1, 0_1))$$

In conformity to the internalist manifesto, Button and Walsh understand determinacy not in terms of what is satisfied in all intended models but in terms of what is proved from every internal-structure. More precisely:

$\varphi(N_1, S_1, 0_1)$ is determinately true if and only if $\forall N_1, S_1, 0_1(\mathsf{PA}_2^{\mathrm{int}}(N_1, S_1, 0_1) \rightarrow \varphi(N_1, S_1, 0_1))$

$\varphi(N_1, S_1, 0_1)$ is determinately false if and only if $\forall N_1, S_1, 0_1(\mathsf{PA}_2^{\mathrm{int}}(N_1, S_1, 0_1) \rightarrow \neg\varphi(N_1, S_1, 0_1))$

Observe that if determinacy is understood in terms of what is proved from every internal-structure, then Corollary 1.2.8.1 establishes that every sentence in arithmetic is determinate. And, finally, since the Theorem amounts to nothing more than the claim that a given formula is provable in the recursive system $D_2$, it bypasses the earlier difficulties with the semantics in Dedekind's Categoricity Theorem. In this sense, Button and Walsh seem to have secured semantic determinacy in a naturalistic-friendly fashion, avoiding mention of models, special faculties of mathematical intuition and unreasonable human cognitive capacities.

Are we done, then? Unfortunately, though an elegant argument in itself, it is not clear whether determinacy should be understood as Button and Walsh advise. Any author is free to use words as they like. But if Button and Walsh aim to capture the same concept as we do, their

definitions seem infelicitous. This is because the concept of determinacy, at least as we have been understanding it here, is linked to the underlying concept of facts of the matter: if a sentence is determinate then there is some fact of the matter about whether it is true or false. As Picollo and Waxman (in press) explain, this means that determinacy is a *metasemantic* concept: it is about what explains that some sentences have the semantic features that they do have. And this, they argue, comes with special explanatory requirements. For it would be unsatisfactory to claim that a sentence just happens to be determinate as a matter of brute fact. On the contrary, every instance of determinacy should be amenable to explanation: if a sentence is determinate, then there must be something that explains why it is so. Nonetheless, understanding determinacy in terms of what is proved from every internal-structure precludes explanation in some cases.

Picollo and Waxman give an example. They note that with respect to a deductive system like $D_2$, the recursively axiomatizable theory $\mathsf{PA}_2^{\text{int}}(N, S, 0)$ has an independent Gödel-Rosser sentence $R$ such that neither it nor its negation is provable from the theory's axioms alone. Suppose for the sake of argument that Corollary 1.2.8.1 establishes semantic determinacy. It follows that $R$ is determinate, so that there must be something that explains why it is so. But how can this be? Since by construction $R$ is an independent sentence, we cannot appeal to facts about provability-in-$D_2$ in our explanation. And it is very difficult to see what other resources are there at our disposal, especially when we are advised to steer clear of semantic notions as per the internalist manifesto. For these reasons, neither first nor second-order internal categoricity seem to improve on Dedekind's original Categoricity Theorem. Let us try something different.

### Reply 3: $\Sigma_1$-Soundness

Recall that according to plenitudinous platonism, mathematical reality is so large that any model of a *consistent* theory is intended in the sense of being true of some part of that reality. Under this picture there does not seem to be a determinate answer to independent sentences. There are intended models corresponding to some portion of reality where the numbers are such that $\mathsf{Gödel}_{\mathsf{PA}}$ is true, and intended models corresponding to a portion where it is false.

However, models of consistent theories can get the wrong facts about consistency itself. By

Gödel's Second Incompleteness Theorem, there is an independent sentence $\mathsf{Con}(\mathsf{PA})$ expressing the theory's own consistency. Both $\mathsf{PA} + \mathsf{Con}(\mathsf{PA})$ and $\mathsf{PA} + \neg\mathsf{Con}(\mathsf{PA})$ are consistent and so have intended models asserting and denying the consistency of $\mathsf{PA}$ respectively. If so, plenitudinous platonism becomes unstable: on the one hand it relies on the consistency of a given theory for its very formulation, but on the other it is committed to indeterminacy regarding whether a theory is in fact consistent. In light of this problem, Clarke-Doane suggests a slight revision of plenitudinous platonism:

> If we replace the consistency requirement with an arithmetic soundness requirement, then only theories that are right about finiteness, consistency, and so on will count as true of their intended subjects. (Clarke-Doane, 2020a, 161)
>
> A still quite radical position which would secure the objectivity of consistency statements would be that any *Pi_1 sound* [sic] theory is true of its intended subject. (Clarke-Doane, 2020b, 2019)

As I understand it, the idea is that all and only the models of $\Sigma_1$-sound extensions of arithmetic are intended.[15] Indeed, this 'light' plenitudinous platonism blocks the original argument from Gödel's First Incompleteness Theorem against semantic determinacy: the theory $\mathsf{PA} + \neg\mathsf{Gödel}_{\mathsf{PA}}$ is $\Sigma_1$-unsound and so its models become unintended under the new criterion. But, of course, it is easily circumvented: it is enough to replace the Gödel-sentence with one of greater complexity. At this point, let us take note of two propositions in connection to a theory's soundness and $\omega$-consistency:[16]

**Theorem 1.2.9** (Kreisel)**.** *There is an $\omega$-consistent extension of PA that proves a false $\Sigma_3$-sentence.*

**Theorem 1.2.10.** *If a theory is $\Sigma_0$-complete and $\omega$-consistent, then it is $\Sigma_1$-sound.*

[15] As Berry (2023, 469 fn. 2) has already observed, Clarke-Doane's use of $\Pi_1$-soundness is probably a typo since any consistent extension of $\mathsf{PA}$, including $\mathsf{PA} + \neg\mathsf{Con}(\mathsf{PA})$, is automatically $\Pi_1$-sound. A theory T is $\Sigma_n$-sound if for any sentence $\varphi \in \Sigma_n$: $T \vdash \varphi \Rightarrow \mathbb{N} \models \varphi$. It is $\Sigma_n$-complete if for any sentence $\varphi \in \Sigma_n$: $\mathbb{N} \models \varphi \Rightarrow T \vdash \varphi$. Similarly for $\Pi_n$-soundness/completeness.

[16] For the proofs see Isaacson (2011, 140–141) and in particular theorems 18 and 19 therein.

We now reason as follows. First, let $\mathsf{T}$ be an $\omega$-consistent theory extending $\mathsf{PA}$ and which proves a false $\Sigma_3$-sentence, as given by Theorem 1.2.9. Since $\mathsf{T}$ is $\omega$-consistent, it is also consistent. Recall that $\mathsf{Q}$ is $\Sigma_1$-complete (Smith, 2013, 79, Theorem 11.5). Since any consistent extension of a $\Sigma_1$-complete theory is itself $\Sigma_1$-complete, it follows that $\mathsf{PA}$ is also $\Sigma_1$-complete. Because $\mathsf{T}$ is a consistent extension of $\mathsf{PA}$, it is also $\Sigma_1$-complete and, therefore, $\Sigma_0$-complete. By Theorem 1.2.10, $\mathsf{T}$ is $\Sigma_1$-sound. Assuming that any model of a $\Sigma_1$-sound extension of arithmetic is intended, it follows that any model of $\mathsf{T}$ is intended as well. Consider then an intended model $\mathfrak{M}_{\mathsf{T}} \models \mathsf{T}$. Since $\mathsf{T}$ proves a false $\Sigma_3$-sentence $\psi$ it follows that $\mathfrak{M}_{\mathsf{T}} \models \psi$. Let $\mathsf{TA}$ be the theory of True Arithmetic, the collection of sentences $\mathrm{Th}(\mathbb{N}) := \{\varphi \text{ is a sentence} : \mathbb{N} \models \varphi\}$. Note that by construction, $\mathsf{TA}$ is $\Sigma_1$-sound. Similar to before, we can consider an intended model $\mathfrak{M}_{\mathsf{TA}} \models \mathsf{TA}$. Moreover, since $\psi$ is a *false* $\Sigma_3$-sentence, it follows that $\mathbb{N} \models \neg\psi$ from where we infer that $\mathsf{TA}$ proves $\neg\psi$. Therefore, $\mathfrak{M}_{\mathsf{TA}} \models \neg\psi$.

The preceding shows that, even assuming Clarke-Doane's light plenitudinous platonism, there are intended models of $\Sigma_1$-sound extensions of arithmetic which decide the sentence $\psi$ in different ways. As a result, there is still an indeterminate sentence.

#### Taking Stock

I have now reviewed several replies to the Argument from Gödel's First Incompleteness Theorem against the semantic determinacy thesis – namely the case from Dedekind's Categoricity Theorem, from Internal Categoricity and from $\Sigma_1$-soundness – and found problems in all three cases. We could imagine further replies to my replies and so push on the dialectic. Indeed, we could also search for an altogether different argument against semantic determinacy. Nonetheless, what I want to stress is that Gödel's First Theorem introduces a difficult and, despite appearances, intricate challenge, however intuitive the thesis may look initially.

## 1.3 Structural Determinacy

We had seen that semantic determinacy is akin to semantic completeness. Arithmetic is semantically determinate if and only if there is no disagreement among intended models with respect

to the truth-value of any sentence in the language. Still, this is not to say that all intended models have the same structure if arithmetic is semantically determinate. These claims can be easily confused so let me introduce new terminology, what I call 'Structural Determinacy':

Intended models are pairwise isomorphic. (Structural Determinacy)

Arithmetic is said to be structurally determinate whenever there is only one intended model up to isomorphism. We already know that semantic and ontological determinacy are largely independent, as per Schema 1. Are semantic and structural determinacy similarly independent? It is easy to see that structural determinacy is not a necessary condition for the semantic component. Suppose that a model of PA is intended if and only if it is elementarily equivalent to $\mathbb{N}$. In this case intended models are all and only the models of True Arithmetic (TA). Since TA is a first-order theory, it has models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ of different cardinalities. These last two models are not isomorphic. But because they are both elementarily equivalent to the standard model, they are still intended models. This means that the pair $\mathfrak{M}_1$ and $\mathfrak{M}_2$ falsifies structural determinacy. Furthermore, since all intended models are elementarily equivalent to each other, there is no disagreement among intended models, proving semantic determinacy. In a slogan: semantic determinacy is about having the right theory, not the right structure.

What about the other direction? Is structural determinacy a sufficient condition? Here, matters are much more interesting. First of all, the distinction between structural determinacy/indeterminacy resembles what other authors call nonalgebraic/algebraic fields. Here is Shapiro on the latter:

> Notice, first, that no one is troubled by the fact that the commutativity of multiplication is independent of the axioms of group theory. This is because, on all accounts, group theory is not about a single structure that is unique up to isomorphism. Rather, group theory is about a class of related structures. The same goes for field theory, topology, and so on. I call such fields "algebraic." The status of independent statements arises (if it does) only in fields like arithmetic, analysis, and

> perhaps set theory. [...] The idea is that each such field is about a single structure, or isomorphism type. For lack of a better term, I call the fields "nonalgebraic." (Shapiro, 1997, 40–41)

Recall further that Shapiro defines realism in truth-value as the view that 'mathematical statements have objective truth-values independent of the minds, languages, conventions, and so forth, of mathematicians'. And the important point is that he thinks that nonalgebraic fields do entail realism in truth-value with respect to their specific subject matter:

> A field is nonalgebraic if it has a single "intended" interpretation among its possible models or, more precisely, if all of its "intended" models are isomorphic (or at least equivalent). This, together with excluded middle (in the object-language theory) yields bivalent truth conditions in the usual manner. Every sentence, including those independent of the axioms, has a truth-value. (Shapiro, 1997, 50)

His reasoning seems to be the following: if a theory is nonalgebraic then all its intended models are pairwise isomorphic or form an isomorphism-type; either way, it follows that all its intended models are elementarily equivalent so that there is no disagreement among them; but this just means realism in truth-value. And, of course, a parallel argument shows that structural determinacy about arithmetic entails semantic determinacy. In fact, the contrapositive entailment is even easier to argue for: if arithmetic is not semantically determinate then, by definition of semantic determinacy, there is a pair of intended models deciding some sentence in opposite ways; as a result, those models cannot be pairwise isomorphic falsifying structural determinacy.

In Chapter 4, I will question the present understanding of semantic determinacy and, as a consequence, this entailment from structural to semantic determinacy. There, I discuss how a unique model of arithmetic can give rise to different incompatible interpretations of a (compositional) truth-predicate. This, I suggest, opens up the possibility of there being only one intended model and still some arithmetical sentence whose truth is not determinate. All these caveats together justify the following schema (where the question mark is self-explanatory):

Schema 2

Semantic Determinacy $\not\Rightarrow$ Structural Determinacy

Structural Determinacy $\stackrel{?}{\Rightarrow}$ Semantic Determinacy

## 1.4 Referential Determinacy

One of the most popular problems about arithmetical determinacy concerns the reference and meaning of our mathematical vocabulary. For recall that according to realism, there exist objects corresponding to the natural numbers. Moreover, if there are such objects as the numbers, it appears plausible that they be conceived as causally inert abstract entities and, in turn, that number-words refer uniquely to the elements of this abstract realm. Under realism then, for each number-term, it is thought that linguistic use singles out only one referent as the term's possible meaning.

The difficulty lies in justifying this view about reference. For recall that according to the popular causal theory of reference, our use of the word 'Fido' manages to refer to Fido the dog, not in virtue of any knowledge of some definite description true only of Fido, but given that our usage is causally connected to usages of the same term by other speakers who, in turn, were similarly causally connected to other speakers, along a causal chain ultimately leading to someone who, by being causally related to Fido itself, was able to dub the canine through a so-called act of 'initial baptism'. Even though the causal theory explains reference to everyday objects, it is barred to the realist. This is because, on the assumption that numbers are acausal, there is no causal link to be had between us speakers and the numbers. In McGee's (1997, 36) amusing phrasing: 'Mathematical objects aren't like Fido, whom you can hold by the collar while you give him a name.'

Perhaps, given this, the realist should weaken her claims about reference just a bit. Perhaps, that is, the realist can maintain that reference to the natural numbers is at least *determinate up to isomorphism*. To explain. We are assuming that linguistic use singles out certain referents and extensions as the possible meaning of our arithmetical vocabulary. Recall that a model of arith-

metic is an *intended model* if and only if it captures the relevant facts of the matter. Let us say that a model of arithmetic is a *referential intended model* if and only if it is a model whose interpretation of the arithmetical vocabulary is one of those possible meanings fixed by linguistic use. Given this, we can specify yet another way in which arithmetic is said to be determinate, what I call 'Referential Determinacy':

Referential intended models are pairwise isomorphic. (Referential Determinacy)

Is the Referential Determinacy thesis easier to argue for? Alas, no. The reason for this is usually credited to Putnam (1980). Putnam stresses that our linguistic-cum-arithmetical practice – which includes our beliefs about the natural numbers and the axioms we take to be true of arithmetic — is first-order regimentable and so subject to the usual limitative results of first-order logic. This means that there are non-standard models of arithmetic of arbitrary cardinalities satisfying the usual constraints we associate with the natural numbers. Finally, this means that there are possible interpretations of our arithmetical vocabulary, what here we are calling referential intended models, of different cardinalities. Some of these cannot be pairwise isomorphic *contra* Referential Determinacy.

Now, what should we say about the relation between semantic determinacy and referential determinacy? I think it would be a mistake to collapse one into the other, as if the two were actually the same. And I think moreover that referential determinacy does not entail semantic determinacy, at least not without further argument. My reasons stem essentially from an extremely intriguing passage by Clarke-Doane who writes:

> Even if we all refer to $\text{sets}_1$ with "sets", so that even undecidable sentences of ZF set theory, like Axiom of Choice (AC), have (determinate) truth-values, the mere existence of rival set-like universes giving intuitively opposite verdicts on such sentences would be enough to undercut the search for the true axioms. [...] This is why (virtually) everyone agrees that if the metaphysical component of set-theoretic pluralism is true, then the question of what set-theoretic axioms are true is miscon-

ceived [...]'. (Clarke-Doane, 2020a, 165–166)[17]

I suggest that we try to imagine an arithmetical analogue to the scenario described in the quote. Suppose that plenitudinous platonism is true so that arithmetical reality comes in clusters of objects. Let us say that a word or thought is intrinsically intentional whenever it intrinsically refers to an object, regardless of the existence of a prior causal link connecting it to a referent.[18] Suppose further that when using arithmetical terms like '2' or 'plus', we manage to refer uniquely and determinately to the objects in and the relations over one and only one such cluster in virtue of those terms being intrinsically referring. In this case, arithmetic is referentially determinate: the referential intended models are just the model corresponding to that particular cluster. Still, is every arithmetical sentence, especially the independent sentences whose truth-value varies from cluster to cluster, determinately true or determinately false? I contend that this is a separate and independent question. After all, it seems that the size of arithmetical reality and, in particular, the remaining existing clusters constitute relevant facts of the matter which should affect our claims to semantic determinacy, regardless whether they are candidates for reference or not: if a sentence changes truth-value across different clusters then it just does not have a determinate truth-value. As I read him, this much is claimed in Clarke-Doane's last remark: 'if the metaphysical component of set-theoretic pluralism is true, then the question of what set-theoretic axioms are true is misconceived'. And if correct, Clarke-Doane's example together with our arithmetical analogue provide a case where referential determinacy is not sufficient for semantic determinacy.

[17] As I noted at the very start, the terminology varies quite drastically when discussing determinacy. As I understand him, Clarke-Doane calls 'objectivism' what I am here calling 'semantic determinacy'. Regarding his own use of 'determinacy' he writes: 'According to the challenge to explain the determinacy of our mathematical beliefs, there is nothing in our practices, or in the world, that could pin down an intended model of set theory. We know, for instance, that if ZFC is consistent, then so is ZFC + Cantor's Continuum Hypothesis (CH) and ZFC + ¬CH. So what, if anything, could make it the case that we are talking about a (class) model in which CH is determinately true or false? There do not seem to be any causal relations between us and sets which could help us tie down reference [...]' (Clarke-Doane, 2020a, 127); this looks closer to what I call 'referential determinacy'.

[18] This intrinsic theory of reference was derided by Putnam as a form of 'magic' and jocosely caricatured in terms of 'some occult rays — call them 'noetic rays' — [that] connect words and thought-signs to their referents' (1981, 51). See also Smith (2003) for more constructive criticism. Let me stress that it is not my intention to defend the view, but just to use it for illustration.

It is noteworthy that this last claim seems to be held by other authors. According to Hamkins, 'the multiverse view is one of higher-order realism — Platonism about universes — and I defend it as a realist position asserting actual existence of the alternative set-theoretic universes' (2012, 417). We have already seen that from this realist view towards multiple models of set-theory, Hamkins directly infers the existence of set-theoretic sentences without determinate truth-values. And what is important now is that questions about language or determinate reference are completely absent from his discussion, suggesting that Hamkins does not consider them relevant to his overall argument *once a plurality of set-theoretic clusters is assumed*.

The other direction — the question whether referential determinacy is necessary to semantic determinacy — is more curious. This is because some authors do interpret Putnam's challenge against referential determinacy as entailing semantic *in*determinacy. For instance, Berry writes:

> For one thing, the most common way of accounting for definite truth values is through definite reference to the natural numbers. [...] if we concede to the Putnamian skeptic that every model of our first-order mathematical beliefs is an equally acceptable precisification of our number concept, then we are forced to conclude that the truth value of some number-theoretic claim is indeterminate. (Berry, 2021, 301)

This is similar to how Field reads Putnam's paper — it is difficult to improve on Field's own exposition so I quote at some length:

> In the first half of his (1980) article, Hilary Putnam gave what I think is a compelling argument against the objectivist position in set theory; he argued that even accepting the component of standard platonism that says that there is a single set-theoretic universe <V,E>, there is nothing in our inferential practice that could determine the truth-value of typical undecidable sentences. [...] In basic outline, the argument is this:

(ia) There is nothing in our inferential practice that could determine that our term 'set' singles out the entire set-theoretic universe V rather than a closed subpart of V.

(ib) Even on the assumption that it singles out the whole set-theoretic universe V, there is nothing in our inferential practice that could determine that our term '$\in$' singles out the membership relation E on V as opposed to some other relation on V that obeys the axioms we have laid down.

(ii) The indeterminacy in (ia) and (ib) is sufficient to leave indeterminate the truth value of typical undecidable sentences of set theory. (Field, [1998] 2001b, 338-339)

I think that their reasoning can be explained in the following way. Earlier I was rather liberal about just what exactly might count as a fact of the matter or an intended model. Berry and Field, however, assume that the relevant facts of the matter include facts about the possible interpretations of our arithmetical vocabulary. That is, they assume that every referential intended model is also an intended model *tout court*. On the extra assumption that there are different referential intended models, some of which think that an independent sentence $\varphi$ is true while others that it is false, it follows that there are intended models disagreeing on the truth-value of $\varphi$. Hence, $\varphi$ is indeterminate. We then have an argument from referential indeterminacy to semantic indeterminacy, which is to say, from semantic determinacy to referential determinacy. This being the case, the arguments we discussed from intuitiveness and Gödel's First Theorem for and against semantic determinacy, together with their respective replies, are also arguments for and against referential determinacy.

Note further that the entailment will complicate some other theories sketched above. In particular, recall that for Maddy (2011, 63–64) realism towards a unique model of set-theory was enough to secure semantic determinacy. Nonetheless, as Field's quote makes clear, Putnam's argument shows that even after assuming a unique universe of mathematical objects we are free to interpret the relations over them in deviant ways so as to end up with models disagreeing on the truth-value of some sentences: different interpretations of the set-theoretic vocabulary

can, *pace* Maddy, lead to semantic indeterminacy in set theory in the same way that different interpretations of the arithmetical vocabulary can lead to semantic indeterminacy in arithmetic.

If correct, the preceding shows that the realist will have a hard time explaining the determinacy of her arithmetical vocabulary and, moreover, that referential determinacy is necessary though not sufficient for semantic determinacy, justifying the following schema:

Schema 3

Semantic Determinacy $\Rightarrow$ Referential Determinacy

Referential Determinacy $\not\Rightarrow$ Semantic Determinacy

## 1.5 Epistemic Determinacy

So much for referential determinacy. Arguably, semantic determinacy will also affect questions of epistemology. *Prima facie*, if a sentence is indeterminate then it is not knowable whether it is (determinately) true or (determinately) false. Picollo and Waxman (in press, 16) seem to recognise this point, writing that 'indeterminacy in this sense [i.e. to say that a claim is determinate is to say that its truth-value is settled by the content-determining facts] may have epistemic implications – if the twin prime conjecture is indeterminate, that plausibly precludes us from knowing whether it is true or false'. Call 'Epistemic Determinacy' the thesis that:

Every sentence in arithmetic is such that it is knowable that it is true or knowable that it is false.

(Epistemic Determinacy)

Now, is epistemic determinacy sufficient or necessary for semantic determinacy? Well, the problem is that the concept of 'knowability' is terribly elusive. Without a more transparent account of its behaviour, it is not clear how knowability, knowledge and semantic determinacy relate. Therefore, at present, I do not have an argument to offer for either direction. I will just note that Picollo and Waxman seem to think that epistemic determinacy is a sufficient condition, while other authors suggest that it is not a necessary one. For example, Warren and Waxman allude to the possibility of unknowable facts of the matter determining the truth-value of

an independent sentence. They write

> Unlike some notions of determinacy that are discussed in philosophy, the notion as we're understanding it here is *semantic* or *conceptual* — it concerns the content of our mathematical notions. In saying this, we mainly mean to rule out the possibility that the relevant notion of indeterminacy is merely epistemic — an expression of ineradicable ignorance or the like. [footnote omitted] To illustrate: claiming that the continuum hypothesis (*CH*) is indeterminate in this sense is incompatible with claiming that there is a *fact of the matter* about whether or not $\aleph_1 = 2^{\aleph_0}$ that we are incapable of discovering. (Warren & Waxman, 2020a, 479)

Still, I argue in Chapter 5 that strict finitism — the view that there are finitely many natural numbers — is buttressed by supposing that some sentences in arithmetic are determinate and unknowable, motivating the claim that semantic and epistemic determinacy are separate theses.

## 1.6 Conclusion

In this chapter I surveyed five distinct albeit interrelated ways of understanding the problem of arithmetical determinacy: what I called semantic, ontological, structural, referential and epistemic determinacy. Even though they enjoy some degree of independence, it is plausible that our commitment to one particular thesis will end up affecting our commitments towards the remaining four. Therefore, it would be naive to characterize the following chapters as dealing with only one or two theses *at most*, in the pretence that they did not bear on how we think about arithmetical determinacy as a whole. Still, at the risk of incurring in a slight mischaracterization, let me try to identify the main themes which we will encounter below. Chapter 2 discusses an argument for referential determinacy in the context of mathematical structuralism. Chapter 3 considers whether special computations describing the performance of supertasks can bolster the semantic determinacy thesis. Chapter 4 questions the relation between structural and semantic determinacy, and proposes a new way of understanding the latter. Chapter 5 studies the epistemic determinacy thesis within strict finitism. Since each chapter is meant to

be self-contained, some of the terminology and themes already discussed will be reintroduced, redefined or redeveloped as appropriate.

# Chapter 2

# Redundancy And Non-Classicality: Two Objections To The Argument From Tennenbaum's Theorem

**Abstract**

One popular version of structuralism holds that intended interpretations of our arithmetical vocabulary should exemplify a common structure: the structure of the standard model of arithmetic. According to Tennenbaum's Theorem, every computable model of Peano Arithmetic is isomorphic to the standard model. And according to the Argument from Tennenbaum's Theorem, this result entails that intended models of arithmetic are isomorphic to the standard model, thus justifying the structuralist insight. Here I present two objections to this reasoning. First, I argue that the Argument trades on an ambiguity in the notion of a computable function which, once spelled out, shows it to be redundant. Second, I argue that the Argument is not sufficient to prove its conclusion on the basis of there being computable models of non-classical arithmetic which do not exemplify the required structure.

## 2.1 Introduction

According to one popular version of structuralism, arithmetic is about a single structure. It is about the structure common to the standard model. Intended interpretations of our arithmetical vocabulary, that is, the interpretations singled out by our linguistic-cum-arithmetical practice should exemplify this structure. However, with a little sophistication, the Compactness Theorem and other limitative results prove that consistent first-order theories of arithmetic have many non-standard models. These constructions raise an important challenge for the structuralist: that of explaining the determinacy of our arithmetical vocabulary, and why we are justified in dismissing non-standard models as constituting unintended interpretations.

Tennenbaum's Theorem states that there is no computable non-standard model of first-order Peano Arithmetic. Several authors have argued on the basis of this result that non-standard models are unintended, on the grounds that arithmetic is computable in any intended model and therefore, by Tennenbaum's Theorem, that all intended models are isomorphic. Let us call this reasoning the *Argument from Tennenbaum's Theorem*. The Argument helps the structuralist when determining the intended models, defining them in terms of a collection of isomorphic models containing the standard model. As a corollary, non-standard models become unintended.

Here, I raise two objections. First, I argue that the Argument trades on an ambiguity in the notion of a computable function which, once spelled out, shows it to be redundant. Second, I argue that the Argument is not sufficient to prove its conclusion on the basis of there being computable models of non-classical arithmetic which do not exemplify the required structure. The chapter is organised as follows. I start by discussing several ways of understanding the idea that arithmetic is about a single structure (section 2.2). Afterwards, I zoom in on a thesis about the determinacy of our arithmetical vocabulary within structuralism, and explain its connection to the Argument from Tennenbaum's Theorem (sections 2.3 and 2.4, respectively). Finally, I raise the two objections (sections 2.5 and 2.6). I conclude that the structuralist cannot avail herself of Tennenbaum's Theorem when explaining the determinacy of our arithmetical vocabulary.

## 2.2 Structuralism

Arithmetic is said to be about a single structure. Even this short slogan is already fraught with ambiguity. There are at least two ways of parsing the structuralist tenet: in terms of a metaphysical and a referential component.

The metaphysical component takes its cue from Benacerraf's (1965) seminal claim that numbers are not objects. By this it is meant that any putative reduction of the natural numbers to other objects like sets — i.e. any list of identity-conditions of the form '$n = o$', for a natural number $n$ and an object $o$ — is spurious. (Note that in this case $n$ is specified solely in arithmetical terms, while $o$ is specified in non-arithmetical terms.) This is justified on the basis of Benacerraf's observation that a reduction of the natural numbers in terms of the finite Zermelo ordinals (with $0 = \emptyset$ and $n + 1 = \{n\}$) and one in terms of the finite von Neumann ordinals (with $0 = \emptyset$ and $n+1 = n \cup \{n\}$) are both admissible.[1] He contends that any attempt to reduce the natural numbers to some other objects is foiled by there being multiple admissible reductions. This is Benacerraf's 'multiple reductions argument'. From this, *eliminative* structuralists infer a distinctive metaphysical thesis: that there is no definite collection of objects which are the natural numbers.

Challenges to the multiple reductions argument are not hard to find: it has been claimed that Benacerraf's argument is question-begging since its success is contingent on already assuming that numbers are not objects (Wetzel, 1989); that it leaves untouched the possibility that numbers are *sui generis* objects not reducible to sets or other mathematical entities (Shapiro, 1997, 79ff); and that it goes wrong in assuming that the Zermelo and von Neumann reductions

[1] We do not need to delve into what counts as an admissible reduction. Benacerraf thinks that a reduction is admissible if and only if it is a model of first-order Peano Arithmetic and it provides an explanation of the application of numbers in counting. These criteria are qualified in his later Benacerraf (1996). On the other hand, Clarke-Doane (2008) argues for a more liberal conception of admissible reductions, suggesting that we should privilege reductions which allow *inter alia* for a more ontologically parsimonious total theory (a theory of the mathematical and the non-mathematical domains) — e.g. on the assumption that we already require an ontology of properties to account for the meaning of predicate-terms, it follows that a reduction of natural numbers in terms of properties is preferable to one in terms of sets given that a total theory only committed to properties is more parsimonious than a total theory committed to properties plus sets.

are both admissible (Clarke-Doane, 2008). Given all this, *non-eliminative* structuralists opt for an equally distinctive metaphysical thesis: that there is a definite collection of objects which are the natural numbers. According to *ante-rem* structuralism in particular, numbers are identified with members ('positions') of the abstract structure that all $\omega$-sequences have in common. An analogy is normally made with the particular-universal distinction: whereas according to eliminative structuralism, arithmetic is about a bunch of particulars — collections of objects all exhibiting a common structure — none of which is more entitled than any other to bear the honorific 'the natural numbers', according to *ante-rem* structuralism, the natural numbers are identified with a universal — an abstract structure — corresponding to what those particulars have in common.

But there is another way of reading the structuralist slogan, namely, as concerning the determinacy of our arithmetical vocabulary. Curiously, this reading also takes the multiple reductions argument as a starting point. Assume that reference is solely fixed by facts about linguistic use (see Dummett ([1973] 1978), Putnam (1977), Shapiro (1991, section 8.3)). If number-words do refer to sets then there should be something about our practice that determines whether they refer to Zermelo or von Neumann ordinals. But on the basis of Benacerraf's observation that both reductions are admissible, it seems that there is nothing in our use of number-words which fixes a unique referent. Similar to before, any attempt to fix a unique referent is bound to be foiled by there being multiple admissible reductions. From this, eliminative structuralists contend that number-words are not singular terms but variables ranging over the elements of admissible structures. Still, under this account, reference candidates do not have to be unruly. That is, the possible values of number-words do not have to be objects from different non-isomorphic models. On the contrary, the eliminative referential thesis goes on to state that our arithmetical vocabulary is determinate in the sense that linguistic-cum-arithmetical practice fixes one particular structure as the interpretation of the vocabulary, namely, the structure of the standard model of arithmetic $\mathbb{N} = \langle \mathbb{N}, 0, S, +, \times \rangle$.[2]

It is unfortunate that the eliminative metaphysical and the eliminative referential theses are

[2] I adopt the convention of letting '$\mathbb{N}$' stand ambiguously for a structure as well as its domain.

prone to conflation. Indeed, Benacerraf introduces them in the same breath apparently exhibiting little caution in making a distinction:

> Why so many interpretations of number theory are possible without any being uniquely singled out becomes obvious: *there is no unique set of objects that are the numbers*. Number theory is the elaboration of the properties of all structures of the order type of the numbers. *The number words do not have single referents*. (Benacerraf, 1965, 70; my emphasis)

But other authors have been more careful:

> Discussing the program of reducing classical mathematics to set theory, Benacerraf asks, What determines what set, if any, the numeral '7' refers to? [...] Benacerraf raises his question as a problem about ontological reduction, but it doesn't really depend on the reductionist framework. The problem is that nothing in our arithmetical thoughts and practices fixes the referents of the numerals more determinately than "up to isomorphism"; hence a numeral doesn't have a determinate referent. (McGee, 1997, 36–37)

This is important for it is not obvious that the two theses stand or fall together. To see why, suppose first that the eliminative metaphysical thesis is false, and replace it with *ante-rem* structuralism. *Prima facie*, on the assumption that *ante-rem* structuralism is correct, number-words simply refer to the abstract structure's members thus disproving the eliminative referential thesis as well. For instance, whilst defending *ante-rem* structuralism, Shapiro (1997, 141) writes that 'the numeral '27', for example, refers to a place in this [abstract] structure. [...] Within arithmetic [...] reference is determinate and scrutable.' Nonetheless, Assadian (2018) has argued that matters are not so clear. His argument relies on the basic model-theoretic fact that any non-trivial permutation on the domain of a model can be used to induce a distinct isomorphic copy. Assadian applies this technique to the abstract structure itself, generating multiple distinct abstract isomorphic copies so that Benacerraf's argument is reintroduced: there is nothing in our

use of number-words which fixes their reference to the members of some abstract structure as opposed to any other. Given this, number-words are better construed as variables ranging over the structures. (Earlier formulations of Assadian's argument can be found in Balaguer (1998, 80–84), Hellman (2001, 193–196), Hellman (2005, 546), and McGee (2005, 151).) Assadian's argument illustrates how the eliminative metaphysical thesis might be false while the eliminative referential thesis is still true.

What about the converse? Can the eliminative referential thesis be false and the eliminative metaphysical thesis true? I am not aware of any clear-cut example. But let me try to advance one. Perhaps the eliminative structuralist may propose that there is no definite collection of objects which are the natural numbers and, still, believe that singular terms refer *arbitrarily* to the elements of some collection of objects — say, the finite von Neumann ordinals.[3] In this case, number-words are singular terms, even though we do not know and cannot know to which objects they refer.[4] Though I am not sure about the tenability of this position. After all, if number-words refer to the finite von Neumann ordinals albeit arbitrarily shouldn't we think that numbers are objects, namely, sets?

Regardless, what is important for my purposes is that these considerations suggest that the eliminative metaphysical and the eliminative referential theses enjoy some degree of independence, so that they can be studied separately. This is already a non-trivial result given that discussions about the Argument from Tennenbaum's Theorem often assume that that argument's success is independent of structuralism's metaphysical commitments. For example, at the start

[3]It is noteworthy that arbitrary reference in this sense rejects the idea that reference is solely fixed by facts about linguistic use as observed in Breckenridge and Magidor (2012, 380). Furthermore, Assadian (2018, 3212) also mentions in passing the idea of conjoining eliminative structuralism with arbitrary reference.

[4]In an extremely intriguing passage Clarke-Doane (2020a, 165–166) writes: 'Even if we all refer to sets$_1$ with "sets", so that even undecidable sentences of ZF set theory, like Axiom of Choice (AC), have (determinate) truth-values, the mere existence of rival set-like universes giving intuitively opposite verdicts on such sentences would be enough to undercut the search for the true axioms. [...] This is why (virtually) everyone agrees that if the metaphysical component of set-theoretic pluralism is true, then the question of what set-theoretic axioms are true is misconceived [...]'. As I understand it, Clarke-Doane sketches a scenario where we are realists towards the set-theoretic multiverse, towards a multitude of different set-theoretic universes, but think at the same time that our set-theoretic vocabulary refers determinately to the members of one and only one such universe. The view described in the main text can be thought of as an arithmetical version of Clarke-Doane's scenario.

of their paper defending the Argument from Tennenbaum's Theorem, Halbach and Horsten write:

> In this paper we take no stand on the ontology of structures. [...] What we say might be compatible with an *ante rem* or an *in rebus* conception of structures or even with a 'nominalist' structuralism where talk about structures is taken as a mere abbreviation for talking about isomorphic models. (Halbach & Horsten, 2005, 174–175, fn. 2)

In the remainder, I will follow suit and solely discuss the eliminative referential thesis (just 'referential thesis' hereafter).

## 2.3 The Referential Thesis

To repeat, the referential thesis states that reference is not unruly but determinate: that our linguistic-cum-arithmetical practice fixes one particular structure as the interpretation of the arithmetical vocabulary, namely, the structure of the standard model of arithmetic $\mathbb{N}$. This idea can be regimented model-theoretically in two steps: first, it is assumed that the theory of first-order Peano Arithmetic (PA) gives a good logical regimentation of our concept of the natural numbers; and second, that talk of a unique structure should be cashed out in terms of isomorphic intended models. By an 'intended model', here, I will understand a model of PA whose interpretation of the vocabulary in the language of arithmetic $\mathcal{L}_{\mathsf{PA}}$ agrees with our linguistic-cum-arithmetical practice. Recall that we are to assume that facts about linguistic use fix certain referents and extensions as the possible meaning of our number-words and arithmetical predicates. An intended model is then a model whose interpretation of the arithmetical vocabulary is one of those possible meanings fixed by use. In this sense, the concept of 'intended model' tries to give a more (model-theoretic) tractable characterization of talk of a 'reference candidate for our arithmetical vocabulary'. With this caveat in place, the referential thesis can be regimented in the following way:

**Regimented Referential Thesis:** If a model is intended for arithmetic then it is isomorphic to the standard model. Equivalently, intended models form a collection of pairwise isomorphic models and which contains the standard model.

This seems to get things right: on the one hand, it isolates 'the' standard model as providing an intended interpretation of our arithmetical vocabulary, and eschews non-standard models — models of PA non-isomorphic to the standard model — as wrong candidates for reference; on the other, it accepts the conclusions of Benacerraf's multiple reductions argument in that it would be wrong to suppose that the reference of our arithmetical vocabulary is fixed more determinately than up to isomorphism.

But though an intuitive story to tell about the natural numbers, Putnam (1980) has shown that it is surprisingly difficult to argue for it. The general idea behind Putnam's argument is straightforward. The usual limitative results of first-order logic, like the Compactness Theorem and the Löwenheim-Skolem Theorem, prove that if PA has a model then it has many non-standard models. Since the latter are *models* of arithmetic, they satisfy the usual constraints associated with our concept of the natural numbers, namely, the first-order Peano axioms. It then seems that there is nothing in our linguistic-cum-arithmetical practice, in the axioms we take to be true of the natural numbers, which render these non-standard interpretations unintended. (Indeed, a quick application of Compactness also shows that there are non-standard models of True Arithmetic, that is, the collection $\{\varphi \text{ is a sentence} : \mathbb{N} \models \varphi\}$.) Hence, there are intended non-standard models, *contra* the referential thesis.[5]

[5]As to the aforementioned thesis that our concept of the natural numbers is best regimented in terms of the first-order Peano axioms there is space for, if not disagreement, then at least qualification. Indeed, Relaford-Doyle & Núñez (2018) have surveyed the reasoning-processes of subjects engaged in proving basic theorems of number theory. Their research suggests that adults not trained in mathematical induction may often rely on conceptualizations of the natural numbers inconsistent with more standard conceptualizations and, in particular, inconsistent with the principle that the natural numbers are formed through iterated application of the successor function from smaller units, so that 'concepts of natural number — even among college-educated adults — are *not* best characterized by the [Dedekind-]Peano axioms' (2018, 249). Furthermore, different foundational accounts in the philosophy of mathematics might presuppose axiomatizations of arithmetic different from the Peano axioms (an example is provided in Chapter 5 in the context of strict finitism). Still, as long as the relevant concept of the natural numbers is captured in terms of a first-order theory, and as long as that theory admits infinite models, the limitative results of first-order logic will ensure the existence of non-standard models so that Putnam's argument can be reintroduced.

One reply to this argument maintains that the natural numbers obey a principle strictly stronger than the first-order Induction Schema, to wit: the second-order Induction Axiom. By Dedekind's Categoricity Theorem, the strengthened induction axiom together with the usual axioms for zero and successor suffice to characterize the standard model up to isomorphism. Therefore, if we subscribe to the stronger principle, it seems that the axioms we take to be true of the natural numbers will fix a unique isomorphism-type as the intended interpretation of our arithmetical vocabulary vindicating the referential thesis. Of course, this reply is beleaguered by a kaleidoscope of problems too well-known to need any repetition. (See for example Field ([1998] 2001b, Postscript), Parsons (1990), and Weston (1976).)[6]

The Argument from Tennenbaum's Theorem tries to meet the challenge from non-standard models by stressing the role played by computability constraints when determining the intended interpretations.

## 2.4 The Argument from Tennenbaum's Theorem

As the name suggests, the Argument from Tennenbaum's Theorem makes essential use of the following result due to Tennenbaum:

**Theorem 2.4.1** (Tennenbaum's Theorem)**.** *If* $\mathfrak{M} = \langle \mathbb{N}, 0^{\mathfrak{M}}, S^{\mathfrak{M}}, +^{\mathfrak{M}}, \times^{\mathfrak{M}} \rangle$ *is a non-standard model of* PA *then addition* $+^{\mathfrak{M}}$ *is not computable. Equivalently, if* $\mathfrak{M}$ *is a model of* PA *and* $+^{\mathfrak{M}}$ *is computable then* $\mathfrak{M}$ *is isomorphic to the standard model* $\mathbb{N}$*. (For a proof see Kaye, 2011)*

The Argument rests on one further assumption. When learning arithmetic we come to recognise that we can lay out computable procedures to calculate sums. This seems an essential feature of our linguistic-cum-arithmetical practice and, as a consequence, one that should help fix the intended models. As Halbach and Horsten (2005, 177) write: 'numbers are something we can calculate with; if we cannot calculate with objects, then they are not numbers'. Also Quinon and Zdanowski (2011, 3): 'if a given model for arithmetic is intended, that is reflecting

[6] For elaboration on this point I refer the reader to Chapter 1 and, in particular, sections 1.2 and 1.4.

our intuitions about the natural numbers, it should be such that one can perform basic arithmetical operations (addition and multiplication) on elements of this model (numbers from this model)'. It follows that having a computable addition function is a necessary condition for a model to be intended. Models without computable addition simply do not agree with our use of number-words.[7]

From the requirement that intended models have a computable addition function together with Tennenbaum's Theorem it follows, as promised, that all intended models of arithmetic are isomorphic to the standard model. In more schematic form, the Argument from Tennenbaum's Theorem is the following:

> **Premise 1**: If a model of arithmetic is intended then it has a computable addition function.
>
> **Premise 2**: Every model of arithmetic with a computable addition function is isomorphic to the standard model.
>
> **Conclusion**: If a model of arithmetic is intended then it is isomorphic to the standard model.

But note that the **Conclusion** is just the **Regimented Referential Thesis**! The Argument then provides a way to rule out non-standard models as unintended interpretations of our arithmetical vocabulary while warranting the structuralist insight. However, this is not to say that any model isomorphic to the standard model is an intended model. Quite the contrary, there are models isomorphic to the standard model without a computable addition function and, thus, according to the Argument from Tennenbaum's Theorem, unintended. Intended models are a proper subclass of the standard models.[8]

[7] Of course, it does not make sense to say that $+^{\mathfrak{M}}$ is computable if the model's domain $M$ is other than $\mathbb{N}$. In the remainder, I will use expressions like 'models with (or without) a computable addition function' as shorthand for 'models whose domain is $\mathbb{N}$ *and* whose addition function is (or is not) computable'. In this sense, the requirement that intended models should have a computable addition function is to be read as asserting that intended models are models on the usual natural numbers and whose addition function is computable.

[8] This last observation makes clear the importance of formulating the **Regimented Referential Thesis** in terms of a conditional instead of a biconditional: were the thesis the claim that a model of arithmetic is intended *if and only if* it is isomorphic to the standard model, the Argument would only justify the left-to-right direction.

The Argument from Tennenbaum's Theorem has been defended, with different levels of detail and focus, by Horsten (2001), Halbach and Horsten (2005) and Quinon and Zdanowski (2011). (Also of interest in this context, though with a different goal, is Horsten (2012).) In the remainder I present two objections against it.

## 2.5 Redundancy

As I see it, the Argument from Tennenbaum's Theorem is redundant though the redundancy is hidden in the Theorem's proof. Tennenbaum's Theorem is usually proved by contradiction. The idea is to pick some arbitrary non-standard model $\mathfrak{M}$ and some appropriate non-computable set $S$ coded in $\mathfrak{M}$. Then, on the assumption that $+^{\mathfrak{M}}$ is computable, it is proved that $S$ is computable too, contradicting the original claim about the set. Let us say a bit more about this $S$ and why it is not computable.

We should first fix some terminology. Recall that two disjoint sets $A, B \subseteq \mathbb{N}$ are said to be computably inseparable if there is no computable set $S \subseteq \mathbb{N}$ such that $A \subseteq S$ and $B \cap S = \emptyset$. It is well-known that computably enumerable computably inseparable sets exist (see for instance Cutland, 1980, 133). Furthermore, for a non-standard model $\mathfrak{M}$ of PA the standard system of sets, written $SSy(\mathfrak{M})$, is the set of all subsets in $\mathbb{N}$ which are coded in $\mathfrak{M}$ by a formula $\varphi(m, x)$, with $m$ an element from its domain $D$. More precisely, let $p(x)$ abbreviate some $\Sigma_1$-formula representing a function enumerating the (standard) primes. So that $p(0) = 2, p(1) = 3, ...$, with $p(n)$ corresponding to the $n$th prime. We then define $SSy(\mathfrak{M}) = \{A \subseteq \mathbb{N} : \exists m \in D$ such that $A = \{a \in \mathbb{N} : \mathfrak{M} \models \exists y(p(\overline{a}) \times y = m)\}\}$. That is, more informally, a set $A$ belongs to the standard system of sets if and only if there is some $m$, such that for every $a \in A$, the $a$th-prime divides $m$ in $\mathfrak{M}$.

The existence of the non-computable set $S$, as required in the proof, is guaranteed by the following Theorem:

**Theorem 2.5.1.** If $\mathfrak{M}$ is a non-standard model of PA then there is a non-computable set $S \in SSy(\mathfrak{M})$.

*Proof Idea.* (Given the scope of the paper, a sketch of the proof will suffice.) Let $A$ and $B$ be computably enumerable computably inseparable sets and let $\exists y\alpha(y, x)$ and $\exists y\beta(y, x)$ be $\Sigma_1$-formulas representing these sets, where $\alpha$ and $\beta$ are $\Delta_0$-formulas. Since $A$ and $B$ are disjoint, it follows that $\mathbb{N} \models \forall a, b, x < \overline{n}\ \neg(\alpha(a, x) \wedge \beta(b, x))$ for every $n \in \mathbb{N}$. By absoluteness of $\Delta_0$-formulas it follows that $\mathfrak{M} \models \forall a, b, x < \overline{n}\ \neg(\alpha(a, x) \wedge \beta(b, x))$ for every $n$. According to Robinson's Overspill Theorem, if $\theta(x, \overrightarrow{m})$ is an open formula with parameters $\overrightarrow{m} \in M$ and if $\mathfrak{M} \models \theta(\overline{n}, \overrightarrow{m})$ for all standard $n$, then there is some non-standard $c \in M$ such that $\mathfrak{M} \models \theta(c, \overrightarrow{m})$. Hence, by the Overspill Theorem, there is a non-standard $c \in M$ such that $\mathfrak{M} \models \forall a, b, x < c\ \neg(\alpha(a, x) \wedge \beta(b, x))$. Define $S = \{s \in \mathbb{N} : \mathfrak{M} \models \exists a < c\ \alpha(a, \bar{s})\}$. It follows that $A \subseteq S$ and $B \cap S = \emptyset$. Since $A$ and $B$ are computably inseparable, $S$ is non-computable. (See Kaye (2011) for details.) □

Regardless, there is a sense in which $S$ is computable due to the following result:

**Theorem 2.5.2.** There is a Turing machine programme $e$ such that for any infinite sequence of natural numbers $\langle n_0, n_1, n_2, ...\rangle$ there is a model $\mathfrak{M} \models \mathsf{PA}$ such that $e$ enumerates inside $\mathfrak{M}$ a sequence extending it.

This is Hamkins's (2018, 13) Corollary 16(1), a corollary to an earlier result by Woodin (2011). (Hamkins in (2018, 10) discusses the authorship of this and related results.) For any function $f : \mathbb{N} \to \mathbb{N}$, Theorem 2.5.2 guarantees the existence of a particular algorithm coded by a (standard) number $e$ and some model $\mathfrak{M}$ such that the latter thinks that $e$ outputs $\langle f(0), f(1), f(2), ...\rangle$. In $\mathfrak{M}$, programme $e$ halts and agrees with $f$ on standard input. As Hamkins (2018, 13) writes: 'Corollary 16 shows the sense in which every function on the natural numbers can become computable, and indeed, computed always by the same universal program, if only the program is run inside the right model of arithmetic'. As a particular instance of Theorem 2.5.2, there is a non-standard model $\mathfrak{M}$ where the standard output of machine $e$ coincides with the characteristic function for $S$. If this is the case then, in that model, $e$ is a decision procedure for $S$ so that the set is computable after all! Though this seems to contradict Theorem 2.5.1 such an impression is easily cleared up.

We need to distinguish between two senses of 'computable function'. First, let us say that a function $f : \mathbb{N} \to \mathbb{N}$ is said to be *externally computable* if it is computed by a Turing machine programme $e$ inside the standard model. In this case the programme's output inside $\mathbb{N}$ agrees with $f$, meaning that for each $n$ the model $\mathbb{N}$ satisfies a formula stating the existence of a computation history, under programme $e$ and input $n$, which halts and outputs $f(n)$. On the other hand, $f$ is said to be *internally computable* if it is computed by a Turing machine programme $e$ inside a non-standard model. In this new case, machine $e$ might, for some inputs, perform a non-standard number of steps in its computation history before halting. The set $S$ is computable in the latter internal sense only; in Theorem 2.5.1 it is proved that $S$ is not computable but again in the former external sense.

So far so good. But recall why we started talking about $S$ in the first place. The Argument from Tennenbaum's Theorem assumes that the interpretation given to basic arithmetical operations in a non-standard model of PA are non-computable as given by Tennenbaum's Theorem. And the proof of Tennenbaum's Theorem makes use of the fact that the set $S$ is not computable. But Theorem 2.5.2 makes clear that when applying Tennenbaum's Theorem we are working with a privileged concept of 'computability' whereby a function/set is computable in the Theorem's sense if and only if it is externally computable. And if this is the notion in play when discussing the Argument, we should rewrite the relevant premise in a more conspicuous form, to wit:

> **Premise 2**$^*$: Every model of arithmetic with an *externally* computable addition function is isomorphic to the standard model.

Now, what about **Premise 1**? If it employs a different notion of 'computability' then the Argument equivocates in articulating the *computable* models of arithmetic: what is understood by computability changes throughout the argument. This means that this premise too should be understood as claiming something like:

> **Premise 1**$^*$: If a model of arithmetic is intended then it has an *externally* computable addition function.

Yet, to say that intended models are to be determined relative to how a particular machine performs inside the standard model as per **Premise 1**$^*$ is already to say that many decision procedures running inside non-standard models are not sufficient to determine the intended models. For consider some model of arithmetic $\mathfrak{M}$, isomorphic to the standard model, with domain $\mathbb{N}$, and whose addition function $+^{\mathfrak{M}} : \mathbb{N} \to \mathbb{N}$ is not externally computable. By Theorem 2.5.2, the function is computed inside some non-standard model by some programme $e$. However, by the contrapositive of **Premise 1**$^*$, it follows that $\mathfrak{M}$ is not intended. And this shows that according to **Premise 1**$^*$ an internally computable addition function is not a sufficient condition for a model to be intended.

But why should this be so? Of course, a quick dismissal of internal computations on the basis of them involving non-standard models is not in this context of much argumentative use. This means that the premise must be argued for; and it is very difficult to see what can justify **Premise 1**$^*$ short of a requirement like the following:

> **Requirement 1**: When determining the intended models of arithmetic, assertions of computability or non-computability for a function/set are to be judged *solely* relative to how a machine performs when running inside the standard model.

**Requirement 1** makes clear why in **Premise 1**$^*$ the addition function must be computable in the external sense; and it also entails that even if there is a machine which inside some non-standard model computes the addition function in a non-standard number of steps, this is neither here nor there as far as intended models go. However, in the context of the Argument from Tennenbaum's Theorem, it is suspect. After all, it is precisely this privileged status given to the standard model that is at issue. If you already look askance at the standard model being *the sole* intended, then you should be equally doubtful of an 'intended notion of computability' designed to determine the intended models and which is defined precisely in terms of *that* standard model.

If the preceding is correct, the Argument from Tennenbaum's Theorem replicates a problem found in an earlier argument from Benacerraf (1996) for the same **Conclusion**. There, Be-

nacerraf requires that intended models be those with order-type $\omega$; schematically his argument is:

> **Premise 1'**: If a model of arithmetic is intended then it has order-type $\omega$.
>
> **Premise 2'**: Every model of arithmetic with order-type $\omega$ is isomorphic to the standard model.
>
> **Conclusion**: If a model of arithmetic is intended then it is isomorphic to the standard model.

If intended models are those with order-type $\omega$ then, automatically, non-standard models become unintended. Nonetheless, since only the standard model up to isomorphism has order-type $\omega$ it looks as if the guiding intuition behind **Premise 1'** is nothing but:

> **Requirement 2**: If a model of arithmetic is intended then it is order-isomorphic to the standard model.

And again it is precisely this privileged status afforded to the standard model as per **Requirement 2** that is at issue. This analysis of Benacerraf's argument is shared by Halbach and Horsten (2005, 175) who write that '[...] nonstandard models are simply ruled out by this restriction on the order-type of the model. This requirement, however, is very bold *and, in a sense, it begs the question. It is like requiring that the model should be isomorphic to the standard model of arithmetic*' (my emphasis).

As I understand it, Halbach and Horsten object that in order to justify **Requirement 2** Benacerraf is forced to assume that intended models are isomorphic to the standard model. In this sense, Benacerraf's argument is redundant since its success presupposes extra theoretical assumptions which independently prove the desired conclusion. By parity of reasoning, a similar redundancy charge can be leveled against the Argument from Tennenbaum's Theorem. In order to justify **Requirement 1**, it seems that we are forced to introduce extra theoretical assumptions — namely, that only the standard model up to isomorphism is intended of our

arithmetical vocabulary — which already secure determinacy independently of Tennenbaum's Theorem. (For similar criticism see Button and Smith (2012) and Dean (2014).[9])

Now, I do not wish to claim that our redundancy charge is unassailable. On the contrary, its relevance lies in pushing the burden of proof to the side of the advocate of the Argument from Tennenbaum's Theorem. The advocate now has to show how we can justify **Requirement 1** without thereby rendering the entire Argument redundant. Suffice it to say this looks well-nigh impossible. But even disregarding the redundancy charge, the Argument is not *sufficient* to secure the determinacy of our arithmetical vocabulary as I explain next.

## 2.6 Non-Classicality

The second objection to the Argument from Tennenbaum's Theorem arises from non-classical models of arithmetic in the so-called Logic of Paradox. To explain the problem this time we should first be clear on some of the basic properties of these models.

The Logic of Paradox (LP) is a paraconsistent trivalent logic whose truth-values are the elements in the set $\mathbf{3} := \{\{1\}, \{0\}, \{1, 0\}\}$. Whilst classical logic includes truth-values $\{1\}$ and $\{0\}$ for the True and the False respectively, the intuition behind LP's third truth-value is that $\{1, 0\}$ represents the conjunction of the classical values. It stands for True-*and*-False.

The language of LP is the same as that of classical first-order logic with identity. Terms and well-formed formulas are defined recursively as usual. For an $n$-ary relation $R^n$ and a classical structure $\mathfrak{M} = \langle D, I \rangle$ with domain $D$ and interpretation $I$, let $I^+(R^n)$ and $I^-(R^n)$ be the relation's *extension* and *anti-extension*, that is, the sets of (sequences of) objects which fall and do not fall under $R^n$ in $\mathfrak{M}$ respectively. Classically, $I^-(R^n)$ can be identified quite simply with the set $D^n - I^+(R^n)$. The reason is that in a classical structure $I^+(R^n)$ and $I^-(R^n)$ are exhaustive for the domain and jointly incompatible; i.e. $I^+(R^n) \cup I^-(R^n) = D^n$ and $I^+(R^n) \cap I^-(R^n) =$

[9] It is worth observing, as Dean (2014, 148–149) already does, that the claim 'the function denoted by + is computable' is formalisable in arithmetic and provable in PA. Therefore, every model of arithmetic satisfies a formula expressing the computability of addition, meaning that every standard and non-standard model already thinks that addition is computable.

$\emptyset$. However, since in LP we want to allow for objects both falling and not falling under $R^n$, an LP-structure's interpretation function has to directly specify $I^+(R^n)$ and $I^-(R^n)$. Hence, when defining an LP-structure $\mathfrak{M} = \langle D, I\rangle$, we stipulate that for an $n$-ary relation symbol $R^n$: $I(R^n) = \langle I^+(R^n), I^-(R^n)\rangle$, where $I^+(R^n), I^-(R^n) \subseteq D^n$ and $I^+(R^n) \cup I^-(R^n) = D^n$. Unlike classical logic, we do *not* require that $I^+(R^n) \cap I^-(R^n) = \emptyset$. But we do require that $I^+(=) = \{\langle d, d\rangle : d \in D\}$. Apart from this, an LP-structure is just like a classical structure.

For convenience, let us assume that for each LP-structure $\mathfrak{M}$, the language is augmented with individual constants for every element of the structure's domain. In particular, we let the elements of $D$ be their own names such that $I(d) = d \in D$. When the language is augmented thus, we call it the 'language of $\mathfrak{M}$'. The interpretation $I$ can now be extended to fix the denotation of terms as usual: $I(f^n(\langle t_1, ..., t_n\rangle)) = I(f^n)(\langle I(t_1), ..., I(t_n)\rangle)$. A valuation function $v$ with respect to a structure $\mathfrak{M}$ mapping (closed) formulas in the language of $\mathfrak{M}$ to elements in $\mathbf{3}$ is built recursively in the exact same way as strong Kleene semantics $\mathsf{K}_3$ (I follow the presentation given in Priest, 1997):

$1 \in v(R^n t_1, ..., t_n)$ if and only if $\langle I(t_1), ..., I(t_n)\rangle \in I^+(R^n)$

$0 \in v(R^n t_1, ..., t_n)$ if and only if $\langle I(t_1), ..., I(t_n)\rangle \in I^-(R^n)$

$1 \in v(\neg\varphi)$ if and only if $0 \in v(\varphi)$

$0 \in v(\neg\varphi)$ if and only if $1 \in v(\varphi)$

$1 \in v(\varphi \wedge \psi)$ if and only if $1 \in v(\varphi)$ and $1 \in v(\psi)$

$0 \in v(\varphi \wedge \psi)$ if and only if $0 \in v(\varphi)$ or $0 \in v(\psi)$

$1 \in v(\forall x\varphi(x))$ if and only if for all $d \in D : 1 \in v(\varphi(x/d))$

$0 \in v(\forall x\varphi(x))$ if and only if for some $d \in D : 0 \in v(\varphi(x/d))$

Disjunction, Existential Quantification and Material Conditional are defined in terms of the earlier logical operators as usual. Finally, satisfiability and LP-semantic consequence ($\models_{\mathsf{LP}}$)

are defined as preservation of at least truth; i.e. they are similar to classical first-order satisfiability and semantic consequence with the exception that both $\{1\}$ and $\{1, 0\}$ are designated. Hence, given a structure $\mathfrak{M}$ and valuation $v$ we say that $\mathfrak{M}$ is a model of $\varphi$ (write $\mathfrak{M} \models_{\mathsf{LP}} \varphi$) if and only if $1 \in v(\varphi)$. Moreover, $\varphi$ is a semantic consequence of a set of formulas $\Gamma$ (write $\Gamma \models_{\mathsf{LP}} \varphi$) if and only if for every structure $\mathfrak{M}$ and valuation $v$: $\mathfrak{M} \models_{\mathsf{LP}} \varphi$ whenever for all $\gamma \in \Gamma$: $\mathfrak{M} \models_{\mathsf{LP}} \gamma$.

It is readily seen that LP invalidates the Explosion Principle. For closed atomic formulas $Ra$ and $Rb$, just consider a structure $\mathfrak{M}$ and valuation $v$ such that $v(Ra) = v(\neg Ra) = \{1, 0\}$ and $v(Rb) = 0$. It follows that $Ra, \neg Ra \not\models_{\mathsf{LP}} Rb$.

The importance of LP to our discussion of the Argument from Tennenbaum's Theorem stems from a result essentially stating that in LP truth is preserved through the taking of quotient-structures. To explain. Let $\mathfrak{M} = \langle D, I\rangle$ be an LP-structure and let $\sim$ stand for an equivalence relation on the domain, which is also a congruence relation with respect to the function symbols in the language; i.e. for $1 \leq i \leq n$ and $d_i, e_i \in D$, if $d_i \sim e_i$ for all $i$ then $I(f^n)(\langle d_1, ..., d_n\rangle) \sim I(f^n)(\langle e_1, ..., e_n\rangle)$. The *collapsed structure* $\mathfrak{M}^{\sim} = \langle D^{\sim}, I^{\sim}\rangle$ is obtained from $\mathfrak{M}$ and $\sim$ following a two-step recipe: the equivalences classes defined over the domain of the original structure are treated as the new elements; and the element's properties are determined in terms of the conjunction of the properties had by the classes' elements in the original structure. More precisely, the domain $D^{\sim} = \{[d] : d \in D\}$ is the set of equivalence classes formed by the partition of $D$ under $\sim$ (where, as usual, $[d] = \{e : e \sim d \text{ and } e \in D\}$). For constant symbol $c$: $I^{\sim}(c) = [I(c)]$. For $n$-ary function symbol $f^n$: $I^{\sim}(f)(\langle [d_1], ..., [d_n]\rangle) = [I(f)(\langle d_1, ..., d_n\rangle)]$. For $n$-ary relation symbol $R^n$: $\langle [d_1], ..., [d_n]\rangle$ is in the extension (anti-extension) of $I^{\sim}(R^n)$ if and only if there are $e_1 \sim d_1, ..., e_n \sim d_n$ such that $\langle e_1, ..., e_n\rangle$ is in the extension (anti-extension) of $I(R^n)$.

It turns out that formulas in a collapsed structure always keep the truth-value they had in the original structure even though they might 'gain' an additional one. That is, if a formula is True-and-False in the original structure then it stays True-and-False in the collapsed structure. Nonetheless, if a formula is True in the original structure then it is either True *or* True-and-False in the collapsed structure; and if a formula is False in the original structure then it is either False

*or* True-and-False in the collapsed structure. This is quickly established by an induction on the complexity of formulas (see Priest, 2006, 229–230), and it is known as:

**Lemma 2.6.1** (Collapsing Lemma)**.** *For every formula $\varphi$ in the language of $\mathfrak{M}$, $v(\varphi) \subseteq v^{\sim}(\varphi)$.*

With the preliminaries out of the way, let us consider a classical non-standard model $\mathfrak{M} \models \mathsf{PA}$ elementarily extending $\mathbb{N}$. Observe that classical structures are just special instances of LP-structures where no formulas are assigned $\{1, 0\}$. Indeed, in this case, the truth-conditions for (closed) atomic *and* non-atomic formulas just reduce to the truth-conditions of classical first-order logic so that the value assigned to any sentence in the language of $\mathfrak{M}$ by the classical structure will be identical to that assigned by the LP-structure. Given this, we can simply identify any classical model of PA with an LP-model of PA.[10] In this sense, $\mathfrak{M}$ is also an LP-structure and $\mathfrak{M} \models_{\mathsf{LP}} \mathsf{PA}$.

Let $\sim$ be an equivalence relation defined over the elements in $\mathfrak{M}$ and which is also a congruence relation with respect to the basic arithmetical functions, such that: $x \sim y$ if and only if ($x, y$ are standard and $x = y$) or ($x, y$ are non-standard). That $\sim$ is an equivalence and congruence relation is proved in Priest (1997, 228). Let $\mathfrak{M}^{\sim}$ be the collapsed structure obtained from $\mathfrak{M}$ and $\sim$. What does it 'look like'? The successor-graph will just be a copy of $\mathbb{N}$ with an extra point looping at infinity as depicted in Figure 2.1. (This structure is also discussed in Priest (1997, 229).)

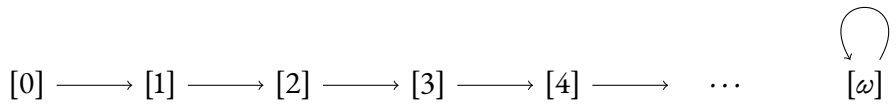



Figure 2.1: The successor-graph of $\mathfrak{M}^{\sim}$.

From the Collapsing Lemma and from the fact that $\mathfrak{M}$ is an LP-model of PA, it follows that $\mathfrak{M}^{\sim}$ is an LP-model of PA. And it is also a non-standard model of PA given that it is not isomorphic to $\mathbb{N}$. (To avoid confusion let me stress that this is *not* to say that $\mathfrak{M}^{\sim}$ is a

[10] Of course, this is a slight abuse of notion. Structures are set-theoretic objects, and both the classical and LP-structures will invariably correspond to different sets given the different ways that their interpretation functions are defined. But nothing heavy will hang on this.

classical model.) Now, it is curious that even though $\mathfrak{M}^{\sim}$ is not isomorphic to $\mathbb{N}$, it is still a computable model. Recall that in a collapsed structure addition of any two classes equals the class that contains the addition of those two classes' elements in the original structure, i.e. $I^{\sim}(+)(\langle [d_1], [d_2]\rangle) = [I(+)(\langle d_1, d_2\rangle)]$. And note further that, in $\mathfrak{M}^{\sim}$, $[\omega]$ is an absorbing element, meaning that $[\omega] +^{\mathfrak{M}^{\sim}} [n] = [n] +^{\mathfrak{M}^{\sim}} [\omega] = [\omega]$. Moreover, addition in $\mathfrak{M}$ coincides with addition in $\mathbb{N}$ for standard elements. Hence:

$$[n] +^{\mathfrak{M}^{\sim}} [m] = \begin{cases} [\omega] \text{ if } [n] = [\omega] \text{ or } [m] = [\omega] \\ [n +^{\mathbb{N}} m] \text{ otherwise} \end{cases}$$

Consider an injection $\pi : D^{\sim} \to \mathbb{N}$ from the domain of $\mathfrak{M}^{\sim}$ to $\mathbb{N}$ such that $\pi([\omega]) = 0$ and $\pi([n]) = n +^{\mathbb{N}} 1$ for $n \in \mathbb{N}$. And let $g : D^{\sim} \times D^{\sim} \to \mathbb{N} \times \mathbb{N}$ be such that $g(\langle [d_1], [d_2]\rangle) = \langle \pi([d_1]), \pi([d_2])\rangle$. Define $+^* : \pi \circ +^{\mathfrak{M}^{\sim}} \circ g^{-1}$ as a function $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$. More informally (but also more wordy): $+^*$ takes (a pair of) two elements in $\mathbb{N}$ has input, adds in $\mathfrak{M}^{\sim}$ the objects that are mapped to those two elements under $g$, and outputs the image of that sum under $\pi$. Since $+^{\mathbb{N}}$ is computable, it is immediate to write a Turing machine programme with the following features: for (the code of) the pair $\langle 0, n\rangle$ or $\langle n, 0\rangle$ the algorithm outputs 0; and for (the code of) the pair $\langle n +^{\mathbb{N}} 1, m +^{\mathbb{N}} 1\rangle$ the algorithm outputs $n +^{\mathbb{N}} m +^{\mathbb{N}} 1$. The programme is an algorithm for $+^*$ which shows that $+^{\mathfrak{M}^{\sim}}$ is computable.[11]

At this point we stumble upon a dialectic in every way identical to the one from the previous section. To risk repetition, the Argument from Tennenbaum's Theorem assumes that the interpretation given to basic arithmetical operations in a non-standard model of PA are non-computable as given by Tennenbaum's Theorem. But the construction of $\mathfrak{M}^{\sim}$ makes clear that when applying Tennenbaum's Theorem we are working with a privileged concept of 'model of PA' which excludes non-classical structures. And if this is the notion in play when discussing

[11] The careful reader might at this point object that, since the domain of $\mathfrak{M}^{\sim}$ is not the natural numbers, it is false that the model is computable. On the contrary, what the result in the main text shows is rather that $\mathfrak{M}^{\sim}$ is isomorphic to a computable model. But since this will not affect the arguments below, I will continue to call $\mathfrak{M}^{\sim}$ computable — always bearing in mind that, in reality, it is this other isomorphic model which is the computable model and that those arguments are actually about this other model.

the Argument, we should rewrite the relevant premise in a more conspicuous form, to wit:

> **Premise 2**$^{**}$: Every *classical* model of arithmetic with a computable addition function is isomorphic to the standard model.

Now, what about **Premise 1** and the **Conclusion?** If they employ a different notion of 'model of PA' then the Argument equivocates in articulating the computable *models* of arithmetic: what is understood by model changes throughout the argument. This means that these clauses too, and so the whole argument, should be understood as claiming:

> **Premise 1**$^{**}$: If a *classical* model of arithmetic is intended then it has a computable addition function.
>
> **Premise 2**$^{**}$: Every *classical* model of arithmetic with a computable addition function is isomorphic to the standard model.
>
> **Conclusion**$^{**}$: If a *classical* model of arithmetic is intended then it is isomorphic to the standard model.

But note that the **Conclusion**$^{**}$ is no longer the **Regimented Referential Thesis**! Or is it? As I have formulated it, the thesis states that if a model is intended for arithmetic then it is isomorphic to the standard model. True, as it stands, it is not clear whether talk of models is supposed to be restricted to the classical ones. Though it is easy to see that it cannot be thus restricted. After all, the **Regimented Referential Thesis** is meant to *regiment* the idea that reference is not unruly but determinate, that our linguistic-cum-arithmetical practice fixes one particular structure as the interpretation of the arithmetical vocabulary. But determinacy is not compatible with rogue non-standard LP-structures like $\mathfrak{M}^{\sim}$ being candidates for reference in the same way that it is not compatible with rogue non-standard classical structures being candidates for reference. Hence, the correct formulation of the **Regimented Referential Thesis** is: if a *classical or non-classical* model is intended for arithmetic then it is isomorphic to the standard model.[12] And this is not the **Conclusion**$^{**}$.

[12] Let me quickly mention one natural objection in this regard. Perhaps it might be said that our mathematical

Bottom line: the Argument from Tennenbaum's Theorem does nothing to exclude $\mathfrak{M}^{\sim}$ as an unintended model.[13] To be clear, it is not my claim that $\mathfrak{M}^{\sim}$ is intended but instead that this (epistemic) possibility has not been ruled out. The Argument is then compatible with assuming that there are intended, albeit non-classical, computable models without the required structure, namely, not isomorphic to the standard model. Finally, it is compatible with the negation of the **Regimented Referential Thesis**, which is to say, it is not sufficient to secure said thesis.

Now, there might be a temptation, pre-theoretically at least, to think that little has to be added to the Argument to address this issue. To think, that is, that although the Argument does not explicitly mention them, the reasons for excluding $\mathfrak{M}^{\sim}$ as an unintended model are rather trivial. I am very sceptical of this suggestion. Let me elaborate.

The first thing to note in this regard is that $\mathfrak{M}^{\sim}$ is not the only computable LP-model of PA. Indeed, the Collapsing Lemma allows us to build *finite* models of arithmetic. Let $n \in \mathbb{N}$ and let $\backsim$ be an equivalence relation defined over the elements in $\mathbb{N}$ and which is also a congruence relation with respect to the basic arithmetical functions, such that: $x \backsim y$ if and only if ($x, y < n$ and $x = y$) or ($x, y \geq n$). Let $\mathbb{N}_n^{\backsim}$ be the collapsed structure obtained from $\mathbb{N}$ and $\backsim$, and whose successor-graph is as depicted in Figure 2.2 (for vividness I assume that $n > 5$).

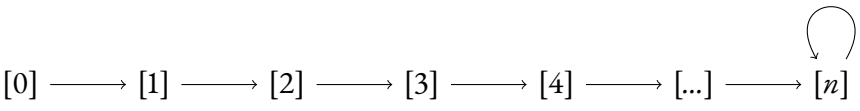



Figure 2.2: The successor-graph of $\mathbb{N}_n^{\backsim}$.

A moment's reflection makes it clear that $\mathbb{N}_n^{\backsim}$ is computable but not of order-type $\omega$. Clearly, this technique can be used to produce infinitely many finite LP-models of PA, one for each

practice already rules out non-classical models given that they invalidate classical rules — e.g. modus ponens or disjunctive syllogism — which seem to be part and parcel of our practice. However, as I try to argue below for the case of modus ponens, perhaps the proper reaction towards certain seemingly paradoxical sentences (like the Liar sentence) will have to involve the rejection of some classical inference rules. This, in turn, suggests that those rules — or, better put, that unrestricted reliance on those rules — are not part of our mathematical practice.

[13] I think it is important to stress that, quite independently of the debate on intended models of arithmetic, it is interesting that there are computable non-standard models of PA in non-classical logic, and that Tennenbaum's theorem itself is not true in this paraconsistent setting.

$n \in \mathbb{N}$. The structuralist who backs the referential thesis will also have to explain why these finite structures are unintended. However, at this point, Putnam's challenge from non-standard models rears its ugly head: since the finite structures are *models* of arithmetic, they satisfy the usual constraints associated with our concept of the natural numbers, namely, the first-order Peano axioms. Sure, some axioms are false in some collapsed models (note that in Figure 2.2 successor is not an injective function) but they are also true. It then seems that there is nothing in our linguistic-cum-arithmetical practice, in the axioms we take to be true of the natural numbers, which renders these non-standard interpretations unintended. As Priest (1994a, 339) writes: 'Putnam's argument appeals, essentially, to the Löwenheim-Skolem Theorem. The Collapsing Lemma that is used to construct the finite interpretations of arithmetic is the ultimate downwards Löwenheim-Skolem theorem: arithmetic has a model of *every* cardinality. The argument from multiplicity [of interpretations] to indeterminacy is therefore reinforced'. Similar remarks apply to $\mathfrak{M}^{\sim}$.

But cannot the finite structures be excluded on the grounds that they are finite? After all, our arithmetical counting practices should preclude an interpretation of arithmetic where the numbers 'run out'. It is not as if we are disposed to count the numbers through 1, 2, 3, ... and suddenly stop at some $n$. Implicit in this reasoning seems to be a form of metasemantic dispositionalism whereby reference is fixed by the speaker's dispositions. Nonetheless, Priest in (1994a, 340–341) and in (1996, 656–657) argues that the reasoning is too facile. As I understand it, his argument is the following. Priest asks us to 'fix $n$ [in $\mathbb{N}_n^{\sim}$] as some incredibly large number, say, a number larger than the number of combinations of fundamental particles in the cosmos, larger than any number that could be sensibly specified in a lifetime, so large that it has no physical meaning or psychological reality' (1994a, 338). Furthermore, we are to imagine two agents A and B counting according to different rules. Agent A will often agree with agent B: they count in the exact same way through the numbers 1, 2, 3, ... They diverge in that B thinks that, at some point, she will always be counting the same number in virtue of identifying $n = n + 1 = n + 2 = \ldots$ Agent A makes no such identification. Note that since $n$ is by construction 'some incredibly large number', our *actual* counting practices can be coherently

described as following either agent/rule A or B. Hence, any evidence that we count more like A rather than B will have to be based on our dispositions to act in *non-actual* scenarios. And it is at this juncture that Kripkenstein's monster jumps out of the bush to claim yet another victim: since rule-following ascriptions can be underdetermined, our intuitions about dispositions to act in non-actual cases are very unreliable. According to Priest, this casts doubt on the idea that we (have a disposition to) count more like A rather than B and, therefore, that our counting practices exclude interpretations like $\mathbb{N}_n^{\frown}$.

Of course, there is a small industry engaged in answering Kripkenstein's challenge (see for instance Warren (2020a) and the references therein) so that Priest's reply is only tentative. Still, for our discussion of the Argument from Tenennbaum's Theorem, what is important is that determining whether the objection to finite structures is successful or not will be a very difficult problem — after all, its success hinges on disputable theoretical assumptions, like that of presupposing that Kripkenstein's challenge to dispositionalism is itself not successful.

The preceding should make us apprehensive of the existence of simple and trivial reasons for excluding LP-structures as unintended. To drive this point home, let us look at one final option. Perhaps, LP-structures are excluded on the basis that they invalidate popular classical inference rules like modus ponens. We should qualify this claim carefully for, in a sense, every LP-structure is actually a model for modus ponens. This is because $\varphi$ is a classical logical truth if and only if it is an LP-logical truth, that is, $\models \varphi \Leftrightarrow \models_{\mathsf{LP}} \varphi$ (for a proof see Priest, 2006, 78). Let $\rightarrow$ be the material conditional with $\varphi \rightarrow \psi$ abbreviating $\neg\varphi \vee \psi$. It follows that $\models_{\mathsf{LP}} (\varphi \wedge (\varphi \rightarrow \psi)) \rightarrow \psi$ for any $\varphi, \psi$. Still, in another sense, LP does invalidate modus ponens given that $\varphi, \varphi \rightarrow \psi \not\models_{\mathsf{LP}} \psi$ for some $\varphi, \psi$. Just let $\varphi, \psi$ be closed atomic formulas and assume that $v(\varphi) = \{1, 0\}$ and $v(\psi) = \{0\}$. But note that $\varphi, \varphi \rightarrow \psi \models_{\mathsf{LP}} \psi \vee (\phi \wedge \neg\phi)$.

Call a connective $\odot$ *detachable* whenever $\varphi, \varphi \odot \psi \models_{\mathsf{LP}} \psi$ for all $\varphi, \psi$. Moreover, call $\odot$ *non-trivially detachable* whenever it is detachable and also is such that $\varphi \odot \psi \not\models_{\mathsf{LP}} \psi$ for some $\varphi, \psi$. Conjunction $\wedge$ is detachable but not non-trivially detachable since $\varphi, \varphi \wedge \psi \models_{\mathsf{LP}} \psi$ but also $\varphi \wedge \psi \models_{\mathsf{LP}} \psi$ for all $\varphi, \psi$. Finally, call a connective *hyperclassical* if it maps $n$-tuples of classical values to classical values.

It is interesting to note that the fact that material conditional is not detachable is essential for the way we have structured our objection from non-classicality. This is for three reasons. First, a non-trivially detachable connective is not definable in LP (Beall, Forster, & Seligman, 2013, 17 Theorem 4.1) Second, the Collapsing Lemma fails for a logic which augments LP with a detachable truth-functional hyperclassical conditional (Ferguson, 2019, 261 Corollary 13.5). Hence, the presence in LP of a non-trivially detachable truth-functional binary connective whose semantics agrees with the classical conditional for classical values is only achieved by augmenting the language; moreover, in that case the Collapsing Lemma cannot be proved. Third and finally, were $\rightarrow$ detachable, the successor-graph of our collapsed models would change considerably: e.g. in $\mathbb{N}_7^{\curvearrowright}$ the element [7] is such that $[7] = S([7])$; therefore, from the formula $S([6]) = S([7]) \rightarrow [6] = [7]$, itself an instance of a PA-axiom, it would be possible to infer $[6] = [7]$ (a similar observation is found in the discussion between Denyer (1995, 573) and Priest (1996, 656)).

What this shows is that if we are able to explain why intended models should validate the modus ponens rule then we can dismiss the problematic LP-structures as unintended. But, again, coming up with such an explanation will be a very tall order. I note in passing that, quite independently of paraconsistent arithmetic, McGee (1985) has famously provided examples of mundane everyday reasoning which he thinks are actual counterexamples to modus ponens in natural language (for discussion Lowe (1987); Piller (1996); Sinnott-Armstrong, Moor, and Fogelin (1986)). Furthermore, observe that a non-detachable material conditional like $\rightarrow$ in LP offers a solution to the semantic paradoxes. It allows for a 'glutty' treatment of the paradoxical sentences whilst blocking triviality. For example, take a Liar sentence $L$ which is True-and-False. This is represented in LP in terms of a valuation $v$ such that $v(L) = \{1, 0\}$. Assume that $v(\varphi) = 0$ for some $\varphi$. Given the definition of $v$, $v(\neg L) = v(\neg L \vee \varphi) = v(L \rightarrow \varphi) = \{1, 0\}$. Since $\rightarrow$ is non-detachable, $\varphi$ cannot be inferred and $L, L \rightarrow \varphi \nvDash_{\mathsf{LP}} \varphi$. Similar considerations have led Goodship (1996) to defend that the proper unified treatment of the semantic paradoxes, including the Liar sentence and the Curry paradox, requires that the principles engendering the paradoxes be regimented in terms of a glutty theory with a non-detachable conditional. This

strategy known as 'Goodship Project' has been the target of comprehensive study (Beall, 2011; Omori, 2016; Priest, 2017). All these references show just how difficult it will be to defend the validity of modus ponens.

Together, these responses serve to reinforce the dialectical inadequacy of the Argument from Tennenbaum's Theorem as a method of proving the referential thesis. Non-standard LP-structures are far from an easy fix, with theoretical resources substantially different from those of the Argument being required to exclude them as unintended models.

## 2.7 Conclusion

By requiring that every intended model be computable, the Argument from Tennenbaum's Theorem tries to characterize the intended models of arithmetic in terms of a collection of isomorphic structures. Here, I have presented two objections to the Argument. I have argued that the Argument is redundant since it privileges a particular concept of computability which makes direct reference to the standard model. In addition, I have explained how the existence of computable non-classical models of arithmetic makes clear that the Argument is not sufficient to prove the structuralist's desired thesis. I conclude that the structuralist cannot avail herself of Tennenbaum's Theorem when explaining the determinacy of our arithmetical vocabulary.

# Chapter 3

# No Semantic Determinacy from Supertask Computations

> The point is that the fact that one can *build a machine to inspect things and tell if they are cats* differentiates cats from cats* if one can be sure 'inspect' and 'tell' refer to inspecting and telling, and it is no easier to say how the reference of *these* words is fixed than to say how the reference of 'cat' is fixed.
>
> Hilary Putnam, *Reason, Truth and History*

**Abstract**

According to the Semantic Determinacy thesis, every sentence in arithmetic is determinately true or determinately false. An interesting idea has it that the thesis can be justified by reference to our experience with certain computational procedures describing supertasks. In this chapter I explain why, contrary to appearances, supertask computations are otiose when trying to secure Semantic Determinacy.

## 3.1 Introduction

For any sentence about the natural numbers, is it always intelligible to ask whether it is true? Is there always some fact of the matter which determines its truth-value? Semantic Determinacy is the thesis that there are always such facts, that every sentence in arithmetic is determinately true or determinately false. Semantic Determinacy has given rise to a burgeoning literature. However, absent from the main debates is a curious though ignored argument suggested by Benacerraf and Putnam ([1964] 1983) and involving computational procedures describing supertasks. (To my knowledge, Warren and Waxman (2020b) is the only criticism of their argument found in the literature.) Benacerraf and Putnam's main insight is surprisingly simple. To explain. Let $\varphi(x)$ be a formula with no unbounded quantifiers, and suppose that a computer checks all instances of the sentence $\forall x\varphi(x)$ by performing each step faster and faster: e.g. it takes half a minute to check $n = 0$, a quarter to check $n = 1$, and so on. After a minute, the computer halts. If it does not find a counterinstance then there is fact of the matter — to wit, the result of the computation — securing the sentence's determinate truth. Relatedly, if the computation finds a counterinstance then there is a fact of the matter securing the sentence's determinate falsity. Either way, the sentence is determinately true or determinately false as per the Semantic Determinacy thesis.

Here, I want to argue against their proposal and explain why supertask computations are otiose when trying to secure Semantic Determinacy. As I see it, Benacerraf and Putnam's argument falls prey to an embarrassment of riches. This is because of familiar model-theoretic results explaining how the existence and properties of a computational procedure are all expressible in the language of arithmetic. It follows that in many cases computability assertions are themselves model-relative: different models of arithmetic give rise to different conceptions of computation which decide one and the same sentence in different ways. For instance, there will be a computational procedure in one model of arithmetic checking all instances of $\forall x\varphi(x)$ without ever finding a counterinstance, and another procedure in some other model which eventually does. And if computability is sufficient for determinacy, as Benacerraf and Putnam wish to claim, the

richness in computational procedures will give rise to sentences both determinately true *and* determinately false. Now, even though I contend that this refutes Benacerraf and Putnam' argument, I also want to show that it is the only successful refutation. We will see below how their argument ramifies widely — *prima facie* it makes use of concepts as diverse as conceivability, possibility, supertasks, and infinitary reasoning — and, thus, exposes itself to objections from multiple corners. Another goal in this chapter will be to explain why those other putative objections do not withstand scrutiny.

The chapter is organised as follows. In section 3.2, I clarify the Semantic Determinacy thesis together with the main terminology in this regard. Afterwards, in section 3.3, I offer a comprehensive analysis ('steelmanning', rather) of Benacerraf and Putnam's proposal: I refine the argument's modal structure considerably and dispose of extraneous assumptions, resulting in gradually refined versions of their original reasoning. Now, an immediate objection to Benacerraf and Putnam is their reliance on disputable assumptions about supertasks and the actual physics of spacetime. This section's upshot will be a strengthening of their argument capable of evading those complaints. I then discuss another interesting objection proposed by Warren and Waxman (2020b), to the extent that the success of Benacerraf and Putnam's argument hinges on an unwarranted use of the $\omega$-rule. Though I am sympathetic to the complaint, in section 3.4 I will argue that the problematic $\omega$-rule is in fact dispensable. Finally, in section 3.5, I present my objection to Benacerraf and Putnam and discuss possible replies. I conclude that supertask computations are not successful in securing Semantic Determinacy.

## 3.2 Semantic Determinacy

Semantic Determinacy is the thesis that every sentence in arithmetic is determinately true or determinately false. By 'arithmetic' I refer to first-order Peano Arithmetic (PA). I also call a sentence 'determinate' whenever there is some fact of the matter determining its truth-value. An analogy is often drawn with the phenomenon of vagueness: if Putnam is a borderline case of tallness then 'Putnam is tall' might be indeterminate, given that there is no relevant fact of

the matter determining a classical truth-value for the sentence. My way of talking about facts is purposefully ambiguous. I am happy to admit that the relevant facts determining a truth-value for 'Putnam is tall' should be, say, facts about the world or about the meaning of predicate-terms or even about our knowledge of those predicates' extensions. And I add a similar addendum about the facts in arithmetic: the relevant facts determining an arithmetical sentence's truth-value might be about the way arithmetical reality is organised, or about the reference of number-terms, among many other options. I am interested in providing a workable definition of determinacy as neutral as possible.

Talk of facts can be best understood model-theoretically in terms of so-called intended models. A model of arithmetic is intended if and only if it captures the relevant arithmetical facts for determinacy. Given this, we can say that a sentence in arithmetic is determinately true if and only if it is true/satisfied in all intended models of arithmetic; similarly, a sentence in arithmetic is determinately false if and only if it is false/falsified in all intended models; and it is indeterminate otherwise. Under this picture, unintended models are seen as logical oddities; an artifact of the expressive limitations of first-order logic. In this sense, the Semantic Determinacy thesis states that:

Every sentence in $\mathcal{L}_{\mathsf{PA}}$ is determinately true or determinately false.

where the 'determinately true/false'-bit is explained in terms of intended models. Of course, the preceding offers only a very quick account of the thesis, though a fuller explanation would take us too far afield — I invite the reader to consult Chapter 1 for a more comprehensive explanation of how I understand Semantic Determinacy.

## 3.3 From Supertask Computations to Semantic Determinacy

In the Introduction to their seminal anthology *Philosophy of Mathematics: Selected Readings*, Benacerraf and Putnam write:

> If we take the stand that "nonconstructive" procedures — i.e., procedures that require us to perform infinitely many operations in a finite time — are conceivable, though not *physically* possible (owing mainly to the existence of a limit to the velocity with which the physical operations can be performed), then we can say that there does "in principle" exist a verification/refutation procedure for number theory. For instance, to "verify" that an equation $P(x, y, z) = 0$ has a solution using the "procedure", we *check each ordered triple x, y, z of integers.* (Of course, this requires working forever, or else completing an infinite series of operations in a finite time.) Similarly, to check a statement of the form $(x)(\exists y)P(x,y) = 0$ (read: 'For every $x$ there is a $y$ such that $P(x,y) = 0$') by the "procedure," we have to substitute 0 for $x$, and then check through $y = 0, 1, 2, \ldots$ until we find a $y_0$ such that $P(0,y_0) = 0$; then we substitute 1 for $x$, and look for a $y_1$ such that $P(1,y_1) = 0$; and so on (again this requires an infinite series of operations). What this shows is: The notion of "truth" in number theory is not a dubious one if the notion of a completed actually infinite series (of, say, definitely specifiable physical operations) is itself not dubious. Since many mathematicians do no share intuitionist doubts about the clarity of the actual infinite, it is understandable that such mathematicians are willing to take the notion of number-theoretic truth as precise. (Benacerraf & Putnam, [1964] 1983, 20)[1]

[1] In a later work Putnam hints at a similar argument writing:

> Now, although many people doubt that the continuum hypothesis has a truth value, everyone believes that the statement '$10^{100} + 1$ is a prime number' has a truth value. Why? 'Because the statement is decidable.' But what does that mean, 'the statement is decidable'? It means that it is *possible* to

As it stands, their argument is very condensed. Benacerraf and Putnam seem to establish an entailment between 'a verification/ refutation procedure for number theory' and the idea that 'the notion of truth in number theory is not dubious', though absent further explication it is not clear how the inference should actually be construed. This is made even more abstruse by reference to concepts — 'conceivability', 'physical possibility', 'the actual infinite' — which are not completely transparent, to say the least. (For discussion on conceivability see for instance Yablo (1993); for physical possibility Chisholm (1967, 412) and Fine (2005); and for the actual infinite Linnebo and Shapiro (2019).) Be that as it may, in this chapter, I am not interested in exegetical faithfulness but instead in retrieving from Benacerraf and Putnam's proposal the best case for Semantic Determinacy. In what follows I will consider (what I take to be) gradually better arguments which do seem implicit in their quote.

### 3.3.1 Benacerraf and Putnam's Argument, a First Attempt

Here is one way of understanding Benacerraf and Putnam's argument. Let $\mathsf{GB}(\mathsf{x})$ be an open formula in $\mathcal{L}_{\mathsf{PA}}$ expressing that 'if $x$ is an even number greater than two then $x$ is the sum of two primes numbers (less than $x$)'. Let also $\mathsf{GB}(\overline{\mathsf{n}})$ be the obvious substitution of the variable $x$ with the numeral for $n$. Goldbach's Conjecture, the statement that every even number greater than two is the sum of two primes, has a very simple form in arithmetic as it is expressible by the $\Pi_1$-sentence $\forall \mathsf{x} \mathsf{GB}(\mathsf{x})$.

It is easy to set up a Turing machine which computes the truth-value of $\mathsf{GB}(\overline{\mathsf{n}})$ in a finite number of discrete steps. But suppose moreover that there is a 'Turing-like machine' which

> try out all the pairs of possible factors and see if any of them 'work'. It means that it is *possible* to decide the statement. Thus, the man who asserts that this statement is decidable, is simply making an assertion of mathematical possibility. Moreover, he believes that just one of the two statements
>
> 3. If all pairs $n, m$ ($n, m < 10^{100} + 1$) were 'tried' by actually computing the product $nm$, then in some case the product would be found to equal $10^{100} + 1$.
> 4. If all pairs $n, m$ ... (same as in (3)), then in no case would the product be found to equal $10^{100} + 1$.
>
> expresses a *necessary* truth, although it may be *physically* impossible to discover which one. (Putnam, 1967, 17–18)

decides in a finite amount of time *every* $\mathsf{GB}(\overline{n})$ by, for example, performing each step faster and faster: it takes half a minute to check $n = 0$, a quarter to check $n = 1$, and so on. The machine is set to output some signal if and only if a counterinstance to the Conjecture has been found. (I will not bother with distinguishing between the formal algorithmic programme that a machine implements and the concrete machine that actually implements it.) After one minute we can check whether the Conjecture has a counterinstance by the presence of the signal or lack thereof.

It is evident that the procedure describes a supertask — i.e. the performance of infinitely many discrete steps in a finite amount of time — and that by running this 'supertask computation' we can decide Goldbach's Conjecture in one minute. To fix on some terminology, I will call this procedure the Goldbach supertask computation. Given this, a preliminary attempt at understanding Benacerraf and Putnam's argument is what I call *The Simple Argument from Supertask Computations*:

**The Simple Argument from Supertask Computations**

(i) Run the Goldbach supertask computation.

(ii) Either:

  (a) The computation never finds a counterinstance;

  (b) Or, the computation finds some counterinstance.

(iii) If (a) then there is a (conjunctive) fact of the matter — to wit, the fact that the machine never finds a counterinstance, together with the relevant metasemantic facts about how certain facts determine a sentence's truth-value — *determining* the Conjecture's truth-value. That is, we should take the fact that the computation never finds a counterinstance as a constraint on our arithmetical practice so that in any intended model each relevant instance of $\mathsf{GB}(x)$ is true and, as a result, that $\forall x \mathsf{GB}(x)$ is true.

(iv) If (b), by parity of reasoning, in any intended model $\forall x \mathsf{GB}(x)$ is false.

(v) In either case, Goldbach's Conjecture is determinately true or determinately false.

The Simple Argument's main idea is that our mathematical practice involving computations, even when these describe a supertask, should play a role in selecting the models contributing to determinacy. In a way, this is nothing new. We already think that computability facts help in fixing the truth of more elementary arithmetical equalities. To give an old example, Bertrand Russell ([1907] 1973, 272) observed that 'the proposition that 2 + 2 = 4 itself strikes us now as obvious [...] But the proposition '2 sheep + 2 sheep = 4 sheep' was probably known to shepherds thousands of years before the proposition 2 + 2 = 4 was discovered; *and when 2 + 2 = 4 was first discovered, it was probably inferred from the case of sheep and other concrete cases* [my emphasis]'. And in the same way that facts about the arithmetic of sheep husbandry play a role in how we judge the truth determinacy of 2 + 2 = 4, computability facts about the Goldbach supertask computation help in fixing the Conjecture's truth-value. The Simple Argument simply generalises our intuitions about the more elementary cases.[2]

An immediate worry with the Simple Argument is the starting assumption, however — that we should run the supertask computation. This is because the coherence of the concept of supertask is threatened by well-known puzzles. To see how, let us alter the Goldbach supertask computation slightly: the machine still computes the truth-value of every $\mathsf{GB}(\overline{n})$ in a finite amount of time and, moreover, successively prints along the cells of a tape a 1 or 0 depending on whether $\mathsf{GB}(\overline{n})$ is true or false, respectively. After one minute we get the result of the Conjecture together with an infinitely long tape of 1's or 0's. It is better if we give a name to this new

[2]As stated, the Simple Argument is about Goldbach's Conjecture only. Nonetheless, Benacerraf and Putnam's original argument is supposed to be an argument for semantic determinacy and, in this sense, it should apply to any sentence in arithmetic. Now, the work of Welch (2008) (see also Button, 2009) shows how we can extend the Simple Argument to secure the determinacy of sentences of greater complexity in the arithmetical hierarchy. Since for my purposes it suffices to consider Goldbach's Conjecture and other $\Pi_1$-sentences, in the remainder I ignore this issue.

programme. Call this procedure the Tape computation. The Tape computation is reminiscent of Thomson's (1954, 5) $\pi$-machine which is set to calculate and print in one minute the decimal expansion of $\pi$. But to such a machine Chihara objected that:

> The difficulty, as I see it, is not insufficiency of time, tape, ink, speed, strength of material, power, and the like, but rather the inconceivability of how the machine could actually finish its super-task. [...] Now when the machine completes its task and shuts itself off, we should be able to look at the tape to see what digit was printed last. But if the machine finishes printing all the digits which constitute the decimal expansion of pi, no digit can be the last digit printed. And how are we to understand this situation? (Chihara, 1965, 80)

By parity of reasoning, the Tape computation seems to face a similar charge. After all, when the Tape Computation completes its task we should be able to look at the tape to see what digit was printed last. Since it seems that no digit can be the last digit printed, it is equally not clear how we should understand the situation.

So what to say in this case? Well, it will be helpful to postpone discussion of the Tape Computation until a bit later (section 3.3.3). But even if this question is left open, note that the charge is aimed at the Tape computation which is essentially different from the Goldbach supertask computation, the one used in Benacerraf and Putnam's argument. And this means that whatever worries we may harbour over the Tape computation, they do not obviously transfer to the Goldbach supertask computation. Regardless, it turns out that deciding Goldbach's Conjecture does not even require the performance of a supertask.

### 3.3.2 Physical Possibility

We can do without supertasks. Indeed, the computation only needs to run inside a Malament-Hogarth spacetime. Very generally the idea is the following. In a Malament-Hogarth spacetime there is a world-line $\gamma$ with infinite elapsed time starting at some event $p$, and a further event $q$ containing $\gamma$ in its past light-cone. We let a machine that successively checks each $\mathsf{GB}(\overline{n})$ run its

computation along $\gamma$. Then, abstracting away from issues due to memory or malfunctioning, the machine will have an infinite amount of proper time to test all instances of the Conjecture and send a signal to $q$ if and only if a counterinstance is found — at any event on world-line $\gamma$, the machine can send a signal to $q$. But in this framework there must also be a second world line $\eth$ from $p$ to $q$ with a finite elapsed time. Therefore a second user can use a finite amount of her proper time to travel along $\eth$, and once in $q$ discover the result of the computation on the basis of the presence or absence of the machine's signal. Since the machine takes an infinite amount of proper time to check the Conjecture, no supertask is involved.[3]

The idea now is to understand the computation deciding Goldbach's Conjecture not as a genuine supertask performing infinitely many steps in a finite amount of time, but instead as a machine checking every $\mathsf{GB}(\overline{\mathsf{n}})$ inside the Malament-Hogarth framework. Let us call it the Goldbach Malament-Hogarth computation (GMH computation, for short). It is clear what the problem is this time: we are saddled with assuming the actuality of the Malament-Hogarth framework. However, ideally, we would like an argument for semantic determinacy which avoided heavy assumptions about the physics of spacetime. In fact, we would like an argument which avoided any assumptions about the physics of spacetime. So can we do better?

When commenting on Benacerraf and Putnam's earlier quote, Warren and Waxman write:

> Benacerraf and Putnam suspected that supertask computation was physically impossible, but more recent work has perhaps called this into question. [...] John Earman and John D. Norton have shown, drawing on work by David Malament and Mark Hogarth, that certain relativistic spacetimes allow computers on certain paths to run through infinitely many individual computations in a finite time from the perspective of an external observer. [...] Without making any detailed claims about what these results do and don't show, we will assume for the sake of argument that supertask computation is physically possible. We mean this to be understood as allowing the *epistemic possibility* that such machines *may one day be*

[3]For discussion see Button (2009), Earman and Norton (1993, 1996), Etesi and Németi (2002), Hogarth (1992, 1994, 2004, 2009) and Manchak and Roberts (2022).

*actual*. (Warren & Waxman, 2020b, 1276–1277)

This is an intriguing passage. Why assume that supertask computation is *physically possible*? And, more importantly, is it relevant to the argument? Well, let us take their suggestion on board and assume that the Malament-Hogarth framework and the GMH computation running inside the framework are both physically possible. In this sense, let us assume that:

At our world: $\Diamond_{\text{Physical}}$ Run the GMH computation ($\Diamond_{\text{P}}$ GMH Computation)

Under the familiar account of modal statements this means that there is a possible world $w$, physically possible from the perspective of our world such that in $w$ we run the GMH computation. Now, in $w$, either the computation never finds a counterinstance or it eventually finds some counterinstance. If the computation never finds a counterinstance, there is a fact of the matter — the fact that the machine never finds a counterinstance — determining the Conjecture's truth-value (do not forget!) in $w$. That is, we should take the fact that the computation never finds a counterinstance as a constraint on our arithmetical practice so that, in $w$, in any intended model each relevant instance of $\mathsf{GB}(x)$ is true and, as a result, $\forall x\mathsf{GB}(x)$ is true. On the other hand, if the computation finds a counterinstance then, by parity of reasoning, in $w$, in any intended model $\forall x\mathsf{GB}(x)$ is false. In either case, in $w$, Goldbach's Conjecture is determinately true or determinately false.

The preceding is obviously the result of applying the Simple Argument from Supertask Computations inside $w$.[4] And from this we conclude that it is *physically possible* that Goldbach's Conjecture is determinately true or determinately false. I abbreviate this by writing:

At our world: $\Diamond_{\text{Physical}}$ Determinacy of Goldbach's Conjecture ($\Diamond_{\text{P}}$D)

This is an interesting result but it is not semantic determinacy. To get more, we have to

[4]Strictly speaking, the original Simple Argument from Supertask Computations appeals to the Goldbach supertask computation and in this new case we are using the GMH computation. Of course, nothing heavy hangs on this.

assume more. Luckily the extra assumptions are unproblematic since widely shared. To explain. It is believed that physical possibility is subsumed under metaphysical possibility in the sense that any physically possible world is a metaphysically possible world. And this means that:

$$\text{At every world: } \Diamond_{\text{Physical}}\ \varphi \rightarrow \Diamond_{\text{Metaphysical}}\ \varphi \qquad (\Diamond_{\text{P}} \rightarrow \Diamond_{\text{M}})$$

From this and from $\Diamond_{\text{P}}$D we get:

$$\text{At our world: } \Diamond_{\text{Metaphysical}} \text{ Determinacy of Goldbach's Conjecture} \qquad (\Diamond_{\text{M}}\text{D})$$

Now, mathematical truths are thought to hold of metaphysical necessity. For example, it is is thought that if a model satisfies some sentence $\varphi$, then necessarily$_{\textit{Metaphysical}}$ that model satisfies $\varphi$. That is, $\mathfrak{M} \models \varphi \Rightarrow \Box_{\text{Metaphysical}}\ \mathfrak{M} \models \varphi$. Moreover, if every model satisfies some sentence $\varphi$, then necessarily$_{\textit{Metaphysical}}$ every model satisfies $\varphi$. And if there is a model satisfying some sentence $\varphi$ while there is another falsifying $\varphi$, then necessarily$_{\textit{Metaphysical}}$ there is a model satisfying $\varphi$ and another falsifying $\varphi$.

A particular instance of this idea is that if at any world Goldbach's Conjecture is true under all intended models or false under all intended models then, at that world, it is *metaphysically necessary* that Goldbach's Conjecture is true under all intended models or false under all intended models. As a consequence, if Goldbach's Conjecture is determinate it is necessarily so. Similarly, if Goldbach's Conjecture is indeterminate it is necessarily so. Hence:

$$\begin{aligned}\text{At our world: } &(\Box_{\text{Metaphysical}} \text{ Determinacy of Goldbach's Conjecture}) \vee \\ &(\Box_{\text{Metaphysical}}\ \neg\text{Determinacy of Goldbach's Conjecture})\end{aligned} \qquad (\Box_{\text{M}}\text{D} \vee \Box_{\text{M}}\neg\text{D})$$

It is now immediate to prove:

$$\text{At our world: Determinacy of Goldbach's Conjecture}$$

The steps are: 1. $\Diamond\varphi$ (assumption); 2. $\Box\varphi \vee \Box\neg\varphi$ (assumption); 3. $\Box\varphi$ (by modal logic, 2 and 3); 4. $\varphi$ (by modal logic, 3 and T-axiom). What this shows is that starting with the phys-

ical possibility of the GMH computation, and adding widely shared modal principles about relativised modalities and arithmetic, yields the Determinacy of Goldbach's Conjecture.[5] This suggests a different reading of Benacerraf and Putnam's argument which I call *The Argument from Physically Possible Supertask Computations*:

**The Argument from Physically Possible Supertask Computations**

(i) Assume $\Diamond_P$ GMH Computation.

(ii) By $\Diamond_P$ GMH Computation + 'The Simple Argument from Supertask Computations' it follows: $\Diamond_P D$.

(iii) By $\Diamond_P D + \Diamond_P \rightarrow \Diamond_M + \Box_M D \vee \Box_M \neg D$ + 'Modal Logic' it follows: Goldbach's Conjecture is determinately true or determinately false.

It is instructive to compare this argument to Benacerraf and Putnam's original quote.

[5] At this point it might be objected that my argument seems to rely on the assumption that the models exist out of necessity. However, given a contingent metaphysics together with common assumptions about the modal properties of sets, if a model's domain contains contingently existing objects then the model itself exists merely contingently — in particular, if some object of its domain does not exist, then the model does neither. And the existence of contingent models might call into doubt the necessity of arithmetical determinacy as per the principle $\Box_M D \vee \Box_M \neg D$.

My solution is to assume that there is, for each contingently existing model, a necessarily existing (i.e. existing in every possible world) isomorphic mathematical model satisfying the same sentences (say, one whose domain is formed entirely by mathematical objects) and, moreover, that for every necessarily existing model $\mathfrak{M}$ it is the case that $\mathfrak{M} \models \varphi \Rightarrow \Box_{\text{Metaphysical}}\ \mathfrak{M} \models \varphi$. This guarantees that if at some world $w$ all the models existing in $w$ satisfy $\varphi$ then at every world $v$ accessible from $w$ all the models existing in $v$ satisfy $\varphi$. For suppose otherwise. Suppose, that is, that all the models in $w$ satisfy $\varphi$ but for some world $v$ accessible from $w$ some model $\mathfrak{M}$ existing in $v$ satisfies $\neg\varphi$. There are now two options. If $\mathfrak{M}$ exists necessarily, then it exists in $w$ where it satisfies $\varphi$. But then in $w$ it is necessary that $\mathfrak{M}$ satisfies $\varphi$, meaning that in $v$ the model satisfies $\varphi$ as well. On the other hand, if $\mathfrak{M}$ exists contingently, then there is a necessarily existing model $\mathfrak{M}'$ in $v$ isomorphic to $\mathfrak{M}$ and satisfying $\neg\varphi$; and by parity of reasoning, we can infer that in $v$ the model $\mathfrak{M}'$ also satisfies $\varphi$. Since in both cases we obtain a contradiction, it follows that at every world $v$ accessible from $w$ all the models existing in $v$ satisfy $\varphi$.

### 3.3.3 The '$\omega$th-step'

Interestingly, for Benacerraf and Putnam, the difficulty with securing semantic determinacy via supertask computations is not so much due to the soundness of the conditional — 'if a verification/refutation procedure for arithmetic is conceivable then arithmetical truth is not dubious' — but instead with the coherence of the antecedent. For if supertask computations are somewhat muddled or suspect then the conditional will not be of much argumentative use. The discussion of Malament-Hogarth spacetime solves this problem: all we need for determinacy is the physical possibility of certain verification/refutation procedures which in point of fact do *not* require the performance of infinitely many discrete steps in a finite amount of time.

However, the problem with supertasks was never about an infinity of steps being performed in a finite amount of time, but rather with us being able to place ourselves 'at the $\omega$th step' of the procedure as it were, at a point immediately after infinitely many steps have been performed. This reading of the problem is not new. As I see it, it is at the heart of the seminal debate between Thomson and Benacerraf.

It is well-known that Thomson (1954)'s main objection to supertasks relies on the supposed incoherence of a lamp which is switched on after half a minute, then switched off after a further quarter, then switched on again after an eighth and so on. After one minute has passed, Thomson thinks that the lamp cannot be on since for every time it was switched on there was a subsequent moment when it was switched off; and it cannot be off either since for every time it was switched off there was a subsequent moment when it was switched on. Yet, Benacerraf (1962) gives a simple diagnosis of the apparent contradiction: the description of the lamp is incomplete; indeed, the description does specify what happens to the lamp before one minute has passed, but those earlier states do not determine nor specify the lamp's state at the one minute mark.

He goes on to give a more general diagnosis of Thomson's paper. According to Benacerraf, Thomson's original objections to supertasks assume that 'the absurdity of performing a super-task [...] must be because there must be an $\omega$th task whose performance is *logically* en-

tailed by the performance of the tasks that properly belong to the super-task. The arguments [by Thomson] all aim at showing that such a task could not have been performed and, hence, that the super-task whose existence would have entailed its performance could not itself have been performed' (Benacerraf, 1962, 774; see also pages 777–778). In a response, Thomson concedes Benacerraf's point and goes on to identify problematic supertasks with the ones involving such an $\omega$th step: 'In general, the idea of an $\omega$-task arises from consideration of a bounded $\omega$-sequence of points on the real line. [...] if it is bounded, the sequence will have a least upper bound and there will be some question to ask about the state of the world or about what happens at some time corresponding to that bound. It seems very obvious that when people have difficulties about the idea of an $\omega$-task being completed the difficulties are about such questions' (Thomson, 1970, 134).[6]

I think this is the right context in which to read Chihara's earlier objection against the $\pi$-machine. Recall that according to Chihara, 'when the machine completes its task we *should* be able to look at the tape to see what digit was printed last'. Let us then imagine an $\omega$th-step corresponding to the first point after the $\pi$-machine has completed its task. Now, as I read him, Chihara seems to assume that if we can specify or describe the machine's steps at every point up to the $\omega$th-step, then we should also be able to specify what happens at the $\omega$th-step itself. And, similar to Thomson, it is the difficulty in doing just this what makes the $\pi$-machine look altogether inconceivable for Chihara. Here, it is irrelevant whether the machine takes only one minute or an infinite amount of time to run the entire computation. As long as there is an $\omega$th-step we are at a loss to know how to specify what happens at that step.

So what should we say about the overall coherence of an $\omega$th-step and, more generally, of a supertask? Here I only have a somewhat disappointing answer to give: as I see it, disagreement about supertasks involving an $\omega$th-step boils down to a difference in fundamental intuitions between the upholder and the objector of supertaks, so that the debate just collapses into a higher-order debate about which side bears the burden of proof when it comes to justifying their personal intuitions. But luckily, I can remain neutral. The Argument from Physically

[6] Also of interest in this context is Russell (1936), George (1993, especially 65–72) and the references therein.

Possible Supertask Computations diverges from Benacerraf and Putnam's quote with respect to how modality is understood. I think this last feature is the key to a stronger argument which evades the issues with the $\omega$th-step, as I now show.

### 3.3.4 Metaphysical Possibility

Before I had assumed that Malament-Hogarth computations were physically possible. Having made that foray into physical possibility, I now want to argue that as far as the determinacy of arithmetical truth is concerned, the physical possibility of the computation is beside the point. Determinacy is invariant under the concrete physical environment where the computation is realized. What matters is that (a) the entailment from computations to truth determinacy holds; and that (b) the relevant computation is possible, not actual, with the possibility at play here not being restricted to physically accessible worlds but instead of a metaphysical variety. Benacerraf and Putnam arguably anticipate this point: recall, in their original quote, the reference to the *in principle* existence of a verification/refutation procedure as being enough to secure the determinacy of number theory; they even assume that supertask computations are not *physically* possible! Putnam (1967) too stresses the physical impossibility but mathematical possibility of procedures deciding a number-theoretic sentence (see footnote 1).

In the Argument from Physically Possible Supertask Computations the lion's share of the work is being done by the *metaphysical* possibility of the determinacy of Goldbach's Conjecture as per $\Diamond_{\mathrm{M}}\mathrm{D}$. It is this principle which, in conjunction with $\Box_{\mathrm{M}}\mathrm{D} \vee \Box_{\mathrm{M}}\neg\mathrm{D}$, entails determinacy *tout court*. But the metaphysical possibility that $\varphi$ is, absent further principles, a strictly weaker assumption than the physical possibility that $\varphi$.[7] If so, the metaphysical possibility of the GMH computation is a strictly weaker assumption than its physical posssibility as encoded in $\Diamond_{\mathrm{P}}$ GMH Computation. Call this weaker assumption:

At our world: $\Diamond_{\text{Metaphysical}}$ Run the GMH computation ($\Diamond_{\mathrm{M}}$ GMH Computation)

[7] By this I mean that (a) the physical possibility of $\varphi$ entails its metaphysical possibility in virtue of $\Diamond_{\mathrm{P}} \rightarrow \Diamond_{\mathrm{M}}$; but that (b) the metaphysical possibility of $\varphi$ does not entail its physical possibility given that there are metaphysically possible worlds not physically possible.

With $\Diamond_M$ GMH Computation we can now reason as before: the Simple Argument from Supertask Computations yields $\Diamond_M D$, without the need of using $\Diamond_P \rightarrow \Diamond_M$; and from $\Diamond_M D$ and $\Box_M D \vee \Box_M \neg D$ we infer the determinacy of Goldbach's Conjecture. What this shows is that the Argument from Physically Possible Computations assumes strictly more than what is required: it assumes $\Diamond_P$ GMH Computation instead of the weaker $\Diamond_M$ GMH Computation. What matters for semantic determinacy is that the computations are metaphysically possible.

The stress on metaphysical possibility makes clear that we do not even need to mention the GMH computation in our argument. Once we are allowed to consider metaphysically possible worlds, instead of the more restrictive physically possible worlds only, we may as well note that there are now other kinds of computations at our disposal that do a better job at securing determinacy. We can explore non-standard order structures for time to build the required computations. For example, I take it to be metaphysically possible for time to have an order-type like $\omega^*$.(Recall, $\omega^*$ is an order-type like $\omega$ but upside down, with a greatest element and no least element.) In particular, there is a metaphysically possible world $w$ where time is order-isomorphic to $(\mathbb{Z}_0^-, <)$, the negative part of the integers with zero.

I add that in $w$ there is a machine that at each time-point $-n-1$ checks $\mathsf{GB}(\overline{n})$, for $n \in \mathbb{N}$.[8] Essentially, the machine has been labouring since the beginning of time (though only metaphorically since in $w$ time doesn't have a beginning). As usual, the machine is set to output some signal if and only if a counterinstance to the Conjecture has been found. In this world, the reader exits at time-point 0 and, therefore, can check if the conjecture is true or false by the presence of the signal or lack thereof. This gives us another example, albeit purposefully extravagant, of a metaphysically possible world with the required decision procedure. Call it the Goldbach Backwards computation.

There are two things to note about the computation running in $w$. First, the computation performs infinitely many discrete steps in an equally infinite amount of time. This means that any objection on the grounds that the computation requires the performance of infinitely many discrete steps in a finite amount of time simply fails to understand the case at hand. Second —

[8] I adopt the convention of letting '$\mathbb{N}$' refer ambiguously to the standard model and its domain.

and much more interestingly — the computation's last step is in this case perfectly intelligible: 'What was the last step of the Goldbach Backwards Computation? Well, it was checking GB(0) at time point −1!'. So define:

At our world: $\Diamond_{\text{Metaphysical}}$ Run the Goldbach Backwards computation

($\Diamond_{\text{M}}$ Backwards Computation)

Putting all our observations together, we are finally in a position to build an Argument for Semantic Determinacy which avoids the issues with supertaks, $\omega$th-steps and the actual physics of spacetime. I call it *The Argument from Metaphysically Possible Supertask Computations*:

**The Argument from Metaphysically Possible Supertask Computations**

(i) Assume $\Diamond_{\text{M}}$ Backwards Computation.

(ii) By $\Diamond_{\text{M}}$ Backwards Computation + 'The Simple Argument from Supertask Computations' it follows: $\Diamond_{\text{M}}$D.

(iii) By $\Diamond_{\text{M}}$D + $\Box_{\text{M}}$D $\vee$ $\Box_{\text{M}}\neg$D + 'Modal Logic' it follows:
Goldbach's Conjecture is determinately true or determinately false.

I find this argument the best of both worlds: on the one hand, it builds on the modal intuitions behind Benacerraf and Putnam's proposal dispensing with the assumptions about the physical possibility or even actuality of a Malament-Hogarth spacetime,[9] and the existence of an

[9]There is space for disagreement. As I noted in footnote 7, there are metaphysically possible worlds not physically possible. This is justified on the basis of conceivability arguments where it is imagined that some physically necessary causal law fails — for sake of argument, let the relevant law be that acceleration is proportional to the applied force. On the assumption that conceivability is (usually) a guide to possibility, we infer the existence of a metaphysically possible world corresponding to the imagined scenario. However, post-Kripke's (1980), we have become sceptical of this kind of argument: the proposed imagined scenario might be one *not* where acceleration fails to be proportional to the applied force, but instead where the alien property 'schacceleration' fails to be proportional to the applied force; or, perhaps, it is one where things *look as if* acceleration is not proportional to the applied force so that the argument shows at most that it is *epistemically* possible that the law fails. (For other arguments defending the metaphysical necessity of causal laws see Shoemaker (1998) and Swoyer (1982).)

$\omega$th-step; on the other hand, by exploring metaphysically possible worlds, it avoids actually performing infinitely many operations in a finite amount of time, dispensing with the problematic concept of supertask as found in Benacerraf and Putnam's argument. Besides, the argument makes clear that Warren and Waxman's earlier proviso — accordingly, 'we will assume for the sake of argument that supertask computation is physically possible' (2020b, 1277) — is also dispensable. Finally, the argument can be formalized in T, the system generated by extending the weakest normal modal logic K with a reflexivity restriction on frames. Since everyone subscribes to T (for discussion see Hale, 2020) and to the necessity of arithmetic, I can simply assume that these modal assumptions are correct.

### 3.3.5 Taking Stock

Taking as a starting point Benacerraf and Putnam's ([1964] 1983, 20) argument that the existence of 'a verification/ refutation procedure for number theory' entails that 'the notion of truth in number theory is not dubious' I have considered successively better refinements of their proposal — The Simple Argument from Supertask Computations, The Argument from Physically Possible Supertask Computations, and The Argument from Metaphysically Possible Supertask Computations — all resting on the idea of a computational procedure akin (though not necessarily identical) to a supertask.

I have argued that if there is a flaw with Benacerraf and Putnam's argument, it does not lie on the concept of supertask, an $\omega$th-step nor on the actual physics of spacetime. Yet Warren and Waxman (2020b) seem to find a different kind of flaw, this time based on the $\omega$-rule. It is to their objection that I turn to next.

---

This opens the way for an argument to the effect that every physical necessity is a metaphysical necessity. If correct, then, were we to assume $\Diamond_M$ Backwards Computation, we would also be assuming the *physical* possibility of the computation so that the Argument from Metaphysically Possible Supertask Computations ends up making assumptions about physical possibility. A proper assessment of this issue would take us too far afield. I will just note that Sidelle (2002) and Fine (2005) have objected against all these arguments for the metaphysical necessity of physical necessity.

## 3.4 Warren and Waxman's Objection

### 3.4.1 The Objection

All three earlier versions of Benacerraf and Putnam's argument make essential use of some decision procedure for Goldbach's Conjecture. Ignoring differences of detail, they all postulate some machine running for a finite or infinite amount of time, and checking every instance of $\mathsf{GB}(x)$ in such a way that, after the machine completes its task, we are able to infer the truth-value of the Conjecture through the presence or absence of a signal indicative that the machine has found a counterinstance. Suppose that the machine has executed its programme and did not send a signal. From this we infer that the machine has accepted every instance of $\mathsf{GB}(x)$. As I am understanding them, Benacerraf and Putnam now take this fact as being *sufficient* for inferring the determinate truth of Goldbach's Conjecture. This reasoning is central in The Simple Argument from Supertask Computations and in the two other arguments that contain it.

However, Warren and Waxman (2020b) remind us that Goldbach's Conjecture is not any collection of $\mathsf{GB}(\overline{n})$'s. It is instead the universal sentence $\forall x\mathsf{GB}(x)$. And this, they argue, is not established by the computation. To see why, note that if each $\mathsf{GB}(\overline{n})$ is true (in the sense of being satisfied in the standard model of arithmetic) then it is already provable due to the $\Sigma_1$-completeness of arithmetic. (The formula $\mathsf{GB}(\overline{n})$ only contains bounded quantifiers which means it is a $\Delta_0$-sentence.) From this and from the soundness of first-order logic, we deduce that each such sentence is true in all models of arithmetic and, moreover, true in all intended models. Hence, each instance is already determinately true regardless of the computation's (possible) results. In a similar spirit, Warren and Waxman write that 'in a precise sense [the computation] gives us nothing that is not already given by standard arithmetical theories such as Peano arithmetic (PA), for each of the individual instances is, if true, already provable in PA. More is required to establish the truth of Goldbach's Conjecture' (2020b, 1278).

Now, as they see it, in order to pass from every instance of $\mathsf{GB}(x)$, which by assumption the computation does establish as true, to the universal sentence $\forall x\mathsf{GB}(x)$ we have to assume extra

inferential resources like the $\omega$-rule:

$$\omega\text{-rule: } \frac{\varphi(0) \quad \varphi(\overline{1}) \quad \varphi(\overline{2}) \quad \varphi(\overline{3}) \quad \ldots}{\forall x \varphi(x)}$$

The $\omega$-rule allows the passage from the infinite collection $\{\varphi(\overline{n}) : n \in \mathbb{N}\}$ to the universal generalization $\forall x \varphi(x)$. But the computation does not justify this step. On the contrary, the $\omega$-rule is an extra assumption required to infer the truth of the universal sentence, whereas the procedure is silent on its validity or invalidity. The procedure might function as an heuristic, a short-cut for finding the truth of each $\mathsf{GB}(\overline{n})$ more quickly, but does not show anything already not established by reasoning solely within PA: each $\mathsf{GB}(\overline{n})$ is already provable in PA if true; and the universal sentence $\forall x \mathsf{GB}(x)$ is inferred only at the cost of extra infinitary reasoning whose introduction is not justified by the computation itself. As a result, the computation does *no* explanatory work:

> The Goldbach computation only "explains" the truth of Goldbach's conjecture if the appropriate instance of the omega rule is valid. And since the computation does nothing to tell us that this instance of the omega rule is valid, the computation itself isn't doing any fundamental explanatory work. The computation merely shows that each of *GB*(0), *GB*(1), *GB*(2), ... are true; but if true, these are each derivable from PA. And this, together with the validity of the omega rule, would in itself explain the truth of Goldbach's conjecture. The supertask computation only enables us to see what follows from PA so as to secure the premises for an omega-inference. With the instances guaranteed by PA, and the omega rule already in hand, there is no explanatory work for the supertask computation to do. So even for Goldbach's conjecture alone, the supertask computation is explanatorily impotent. (Warren & Waxman, 2020b, 1278–1279)

That the machine has accepted every instance of $\mathsf{GB}(x)$ is *not sufficient* to infer the determinate truth of the Conjecture itself.[10] Worse, Warren and Waxman's objection generalizes widely.

[10] Warren and Waxman (2020b, 1279–1280) also consider another approach which on the surface avoids the $\omega$-

To explain. Using the fact that Robinson's Arithmetic Q is also $\Sigma_1$-complete, an induction along the arithmetical hierarchy quickly proves that augmenting Q with the $\omega$-rule collapses the theory into True Arithmetic (TA); more formally:

**Theorem 3.4.1.** Q + $\omega$-rule = TA

In particular, it follows that augmenting PA with the $\omega$-rule also yields TA. The theory of True Arithmetic is just the theory of the standard model: $\mathsf{TA} = \{\varphi \text{ is a sentence} : \mathbb{N} \models \varphi\}$. Therefore, all models of TA have the same first-order theory. As a result, if our arithmetical practice validates the use of the $\omega$-rule then all models of the resultant theory are elementarily equivalent, deciding any sentence of whatever complexity in the exact same manner. *A fortiori,* if our arithmetical practice validates the use of the $\omega$-rule then all *intended* models of the resultant theory are elementarily equivalent proving semantic determinacy. This is reminiscent of the case from full $\mathsf{PA}_2$ where determinacy was also ensured by the theory's categoricity and, consequentially, the elementary equivalence among the theory's models (as in Chapter 1 and section 1.2.3.)

What this shows is that the success of Benacerraf and Putnam's argument depends on infinitary reasoning which, by proving semantic determinacy independently of the relevant computation, once available, makes their entire argument dispensable. Or, as Warren and Waxman write:

> Closing standard arithmetical theories, such as PA, under the omega rule, results in true arithmetic: the complete and non-recursive first-order theory of the standard model of arithmetic. This means that if the omega rule is valid in our arithmetical language, then arithmetic is determinate. The omega rule, together with PA, is doing all the work. Supertask computation seems irrelevant to explaining semantic determinacy. (Warren & Waxman, 2020b, 1279)

rule. The idea this time is that since the computation establishes the truth of each $\mathsf{GB}(\overline{n})$ we can use the Induction Axiom $(\mathsf{GB}(0) \wedge \forall x\ (\mathsf{GB}(x) \rightarrow \mathsf{GB}(S(x)))) \rightarrow \forall x \mathsf{GB}(x)$ to get Goldbach's Conjecture $\forall x \mathsf{GB}(x)$. But this is just more of the same, if not worse. By assumption, the computation establishes each $\mathsf{GB}(\overline{n})$ and so it will be true that $\mathsf{GB}(0) \rightarrow \mathsf{GB}(\overline{1})$, $\mathsf{GB}(\overline{1}) \rightarrow \mathsf{GB}(\overline{2})$, $\mathsf{GB}(\overline{2}) \rightarrow \mathsf{GB}(\overline{3})$, ... But to pass from these instances to the universal $\forall x(\mathsf{GB}(x) \rightarrow \mathsf{GB}(S(x)))$ we still require some form of infinitary reasoning like the $\omega$-rule.

On the whole, I think that Warren and Waxman's objection proceeds roughly in two steps. In the first step, they zoom in on Goldbach's Conjecture and call attention to the fact that the universal sentence is only inferred after the introduction of the $\omega$-rule, rendering the computation which decides the Conjecture dispensable. In the second step, they zoom out and *generalize* stating that the introduction of the $\omega$-rule makes arithmetic automatically determinate, rendering every such computation dispensable.

This is an elegant reasoning even though it is noticeably conditional on two somewhat loose assumptions: that each $\mathsf{GB}(\overline{n})$ is true (recall that throughout we have been assuming that the machine never finds a counterinstance); and that the universal sentence $\forall x \mathsf{GB}(x)$ is only inferred after the introduction of the $\omega$-rule. Nonetheless, we might worry that the objection is weakened if there is a counterinstance to Goldbach's Conjecture — granted, an implausible supposition (on why this is implausible see Baker (2007, 69–71)). Furthermore, there are extensions of arithmetic with restricted forms of the $\omega$-rule which achieve some amount of completeness without reaching the full strengths of True Arithmetic; and this seems particularly troublesome to their generalization step. I use the next subsection to briefly explain why these worries are ill-founded.

### 3.4.2 A Refinement

Warren and Waxman's objection is conditional on there being no counterinstance to Goldbach's Conjecture. But suppose that there is one. Suppose that there is some true $\neg \mathsf{GB}(\overline{n})$. In this case, there is no need for the $\omega$-rule. The claim $\exists x \neg \mathsf{GB}(x)$ is proved from one instance of existential-introduction, a rule presumably already included in the deductive system. And from this we infer the determinate falsity of Goldbach's Conjecture. Though it is important to note that in this case the computation deciding the Conjecture is still dispensable: if $\neg \mathsf{GB}(\overline{n})$ is true then it and $\exists x \neg \mathsf{GB}(x)$ are already provable in PA. Hence, a counterinstance to the Conjecture should not give Warren and Waxman too much pause.

However, their objection is equally conditional on the further assumption — instrumental in the generalization step — that the universal sentence $\forall x \mathsf{GB}(x)$ is only inferred after the

introduction of the $\omega$-rule. Yet, the sentence is also provable in systems strictly weaker than True Arithmetic. One such system is an extension of PA where applications of the $\omega$-rule are bounded by a finite ordinal. For example, let $\mathsf{PA}_0 = \mathsf{PA}$ and suppose $\mathsf{PA}_n$ has already been defined. Take the set of formulas $S_{n+1} = \mathsf{PA}_n \cup \{\forall x \varphi(x) : \mathsf{PA}_n \vdash \varphi(\overline{m}),\ \text{for all } m \in \mathbb{N}\}$. The theory $\mathsf{PA}_{n+1}$ is now obtained by closing the set $S_{n+1}$ under the usual PA-rules (which do not include the $\omega$-rule). Each theory $\mathsf{PA}_n$ is then the set of sentences containing the PA-axioms, and closed under the usual PA-rules and *at most* $n$-many applications of the $\omega$-rule. Assuming that each $\mathsf{GB}(\overline{n})$ is true and therefore provable, it follows that Goldbach's Conjecture is provable in $\mathsf{PA}_1$. But is $\mathsf{PA}_1$ complete? This is answered in the negative:

**Theorem 3.4.2** (Rosser, 1937; Theorem 6)**.** Let $n$ be a finite non-zero ordinal and suppose $\mathsf{PA}_n$ is consistent. Then the formula stating the consistency of $\mathsf{PA}_n$ cannot be proved in $\mathsf{PA}_n$.

Inferring the universal $\forall x \mathsf{GB}(x)$ from the many $\mathsf{GB}(0)$, $\mathsf{GB}(\overline{1})$, $\mathsf{GB}(\overline{2})$, ..., can be done inside $\mathsf{PA}_1$. And augmenting arithmetic with rules restricted in this fashion does not result in the complete theory TA nor does it lead to semantic determinacy. This blocks their generalization step. But again, there is an easy fix: we should read the objection's first step as providing something of a *template*. That is, suppose that we use Benacerraf and Putnam's argument to secure the determinacy of any universal sentence $\varphi$, of arbitrary complexity; now reasoning in a way parallel to the objection's first step, in each case, an extension of arithmetic with $n$-many applications of the $\omega$-rule will be required, rendering the computation which decides $\varphi$ dispensable.[11]

Therefore, arguing against Warren and Waxman on the basis of the two conditional assumptions is not that promising. In the next section I turn to a different objection which seems to be in better standing.

[11] Besides, this is such an easy reply that I wonder whether Warren and Waxman did not intend this slightly different argument from the get-go. In fact, when introducing the $\omega$-rule, they write in a footnote: 'Obviously, for this particular case [the case about Goldbach's Conjecture] we could work with a restricted version of the rule requiring that $\varphi$ contain no unbounded quantifiers' (2020b, 1978, fn.10).

### 3.4.3 Is the $\omega$-rule indispensable?

As I see it, the idea that the $\omega$-rule is indispensable relies on an unwarranted assumption. Suppose that the computation has 'finished' its task without sending a signal or, equivalently, without ever finding a counterinstance. According to the Simple Argument from Supertask Computations, as well as the other two versions, there is now a fact of the matter determining that there is no counterinstance to the Conjecture. And this just means:

$$\text{For any intended model } \mathfrak{M} \text{ and domain } M \text{ it follows: } \text{For every } m \in M : \mathfrak{M} \models \mathsf{GB}(\mathsf{x}/m) \qquad \text{(No Counterinstance)}$$

(Of course, this is only loosely written given that $m$ is not a symbol in the language, meaning that '$\mathsf{GB}(\mathsf{x}/m)$' is not well-formed. Yet, the present terminology eases exposition and, besides, it can be easily fixed by, say, treating each element as its own name or by specifying a variable-assignment.) And I think that something has gone wrong if we are now told that we need the $\omega$-rule to conclude that the Conjecture is true in all these models. On the contrary, the following semantic principle holds in the meta-theory:

$$\mathfrak{M} \models \forall \mathsf{x} \varphi(\mathsf{x}) \text{ if and only if for every } m \in M : \mathfrak{M} \models \varphi(\mathsf{x}/m) \qquad \text{(Generalization)}$$

It quickly follows:

$$\text{For any intended model } \mathfrak{M} \models \forall \mathsf{x} \mathsf{GB}(\mathsf{x}) \qquad \text{(Determinacy)}$$

This means that Goldbach's Conjecture is determinately true. We have reached this conclusion without the $\omega$-rule, but instead via one application of Generalization already available in the meta-theory. Yet, this is exactly what Warren and Waxman's objection was supposed to block. So what has gone wrong? I think that the clue is contained in the following passage:

> We know, by familiarity with the mechanism, that the Goldbach computation proceeds by checking each instance in succession. Consequently, the computer's fail-

> ure to halt gives us reason to accept each of the infinitely many claims — $GB(0)$, $GB(1)$, $GB(2)$, ..., $GB(n)$, ... — and so on, for each standard numeral "n". (Warren & Waxman, 2020b, 1278)

Observe their restriction on standard numerals. It seems that according to Warren and Waxman, if the relevant computation never finds a counterinstance then this is at most evidence that it is determinately true that there is no *standard* counterinstance to the Conjecture. In our terminology this means that:

> For any intended model $\mathfrak{M}$ and domain $M$ it follows:
>
> For every *standard* $m \in M : \mathfrak{M} \models \mathsf{GB}(\mathsf{x}/m)$ (No Standard Counterinstance)

This is strictly weaker than No Counterinstance. And more importantly, if we are *only* allowed to use No Standard Counterinstance then, even with Generalization in place, we still cannot conclude Determinacy. This is because, for all we know, some intended models might be non-standard and contain counterinstances to the Conjecture only in their non-standard part. If so, it is clear the need for extra inferential rules like the $\omega$-rule in order to derive Determinacy.

Now, in the previous section we saw how Warren and Waxman's objection was conditional on two assumptions. We now see that the success of their objection hinges on one further idea: that if the computation never finds a counterinstance this is *at most* evidence for No Standard Counterinstance and *not* for No Counterinstance. And the problem is that this has not been argued for. Indeed, the entailment from the computation never finding a counterinstance to No Counterinstance just generalizes our intuitions about how sentences come to be determinate in particular cases. This means that the burden of proof rests with Warren and Waxman. They have to explain why the failure in finding a counterinstance does not entail the stronger principle. As a result, absent further argumentation, Benacerraf and Putnam's original argument still stands.

### 3.4.4 Taking Stock

Warren and Waxman (2020b) present a dilemma to Benacerraf and Putnam: either infinitary reasoning is introduced, securing semantic determinacy at the cost of making every computation redundant, or Benacerraf and Putnam are debarred from introducing infinitary reasoning so that their argument fails to prove its conclusion. I have identified three *prima facie* disputable assumptions behind Warren and Waxman's objection: that there is no counterinstance to the Conjecture; that the universal sentence $\forall x \mathsf{GB}(x)$ is only inferred after the introduction of the $\omega$-rule; and that failure of the computation in finding a counterinstance is not evidence for No Counterinstance. While the first two should in the end pose no difficulty, it is not clear what stops us from rejecting the third assumption. And if we do reject it, it seems that Warren and Waxman's objection does no longer go through. In the next section I present a new objection to Benacerraf and Putnam's argument.

## 3.5 Model-Relative Computability

As I see it, the flaw in Benacerraf and Putnam's argument lies in thinking about a computational procedure as something 'existing in a vacuum', detached from the particular model of arithmetic inside which it is 'running'. For example, they write that 'to verify that an equation $P(x, y, z) = 0$ has a solution using the procedure, we check each ordered triple $x, y, z$ of integers'. But the computational procedure which searches for a solution of $P(x, y, z) = 0$ is computing on integer numbers or, more precisely, on the natural numbers coding the integers, and these as well as the relations holding among them are defined only relative to a background model of arithmetic. More generally, computability assertions are themselves model-relative. The rules of a Turing programme as well as the properties of a computation according to that programme are all expressible via coding. Assertions of the form 'if the reading head is in state $q_1$ and scans symbol 1 on the tape then print 0 and move one cell to the right and change to state $q_2$' and 'the programme halts on input $x$ after $t$ steps with output $y$' are representable in the language of arithmetic. It follows that the steps carried out by a decision procedure for Goldbach's Conjec-

ture on a given input — i.e. the procedure's computation history on that input — together with the result of the computation are also expressible by formulas in arithmetic. It is then natural to ask after these formulas' truth-value inside different models of arithmetic: more informally, what is the computation's output when running inside a given model? And the important point is that computability assertions can change drastically across different models. Let me give a visual of how a programme can run inside a model.

### 3.5.1 A New Decision Procedure For Goldbach's Conjecture

Throughout, Goldbach's Conjecture has served as a case study of how supertask computations might come to secure semantic determinacy. It is striking that nothing in Benacerraf and Putnam's argument requires that the programme searching for a counterinstance to the Conjecture be running inside the standard model $\mathbb{N}$. This is not a mere fluke. On the contrary, given their overall dialectic, demanding that determinacy be fixed by the results of a decision procedure computing *solely* on the standard natural numbers is an unstable position. A requirement of this sort would isolate the standard model as the privileged structure in which to evaluate determinacy. If so, we could take this as evidence that the standard model is the only intended model and, as before, secure semantic determinacy — in this case, their argument becomes redundant. Hence, as far as their argument goes, the decision procedure for Goldbach's Conjecture may be running inside a non-standard (countable) model $\mathfrak{M}$ computing on non-standard numbers.

Recall that $\mathsf{GB}(\mathsf{x})$ abbreviates a formula in $\mathcal{L}_{\mathsf{PA}}$ expressing that 'if $x$ is an even number greater than two then $x$ is the sum of two primes numbers (less than $x$)'. It is clear that $\mathsf{GB}(\mathsf{x})$ is a decidable predicate and so its characteristic function is computed by some Turing programme $e$. Kleene's Normal Form Theorem teaches us that the statement '$y$ codes a halting computation history of the function with index $e$ on input $x$' is representable by Kleene's T-predicate $\mathsf{T}(\mathsf{e},\mathsf{x},\mathsf{y})$. Additionally, there is a formula $\mathsf{U}(\mathsf{y})$ expressing the output of $y$. (For a statement and proof of the Normal Form Theorem see for example Cutland, 1980, chapter 5.) Since these properties are representable in arithmetic we can ask after the results of programme $e$ when computing on the numbers in $\mathfrak{M}$. A bit more loosely, we can ask whether $e$ ever finds a coun-

terinstance when running in $\mathfrak{M}$. And the claim that, for every number $m \in M$, there is some computation under $e$ which does not think that $m$ is a counterinstance to Goldbach's Conjecture is the statement:

$$\mathfrak{M} \models \forall x \exists y(\mathsf{T}(\overline{e},x,y) \wedge \mathsf{U}(y)=\overline{1}) \qquad (1\text{—}\mathfrak{M})$$

That is, for any element $m$ in the model's domain $M$, there is a halting sequence of steps $y$ according to the Turing programme $e$ on input $m$ and which prints '$\mathsf{GB}(x)$ is true of $m$', where this last fact is represented by $e$ outputting 1. On the other hand, that for some number $m \in M$, there is a computation under $e$ which thinks that $m$ is a counterinstance to Goldbach's Conjecture is the statement:

$$\mathfrak{M} \models \exists x \exists y(\mathsf{T}(\overline{e},x,y) \wedge \mathsf{U}(y)=0) \qquad (0\text{—}\mathfrak{M})$$

Since by assumption $\mathfrak{M}$ is a non-standard model, it follows that some witnesses for $x$ in $1\text{—}\mathfrak{M}$ or in $0\text{—}\mathfrak{M}$ will be non-standard elements in $M$. What's more, in that case, some witnesses for $y$ might also be non-standard, meaning that the programme might carry out non-standard many steps before reaching a halting state.

Now, suppose that $1\text{—}\mathfrak{M}$ is the case. We can consider a metaphysically possible world $w$ where a machine runs programme $e$ inside the model $\mathfrak{M}$. That is, for each element of the domain $m \in M$, the machine executes the steps encoded by the element $n$ for programme $e$, where $n$ is such that $\mathfrak{M} \models \mathsf{T}(\overline{e},m,n) \wedge \mathsf{U}(n)=\overline{1}$. This is, the element $n$ is the value that $y$ takes whenever $x$ takes the value $m$ in $1\text{—}\mathfrak{M}$. In essence, the machine goes through all the numbers in $M$ and checks whether each number is a counterinstance to the Conjecture using programme $e$. As usual, it is designed to send some signal if and only if a counterinstance has been found. How is time structured in this possible world $w$? Well, suppose that for each time-point in $w$ the machine executes one step. Then, in this world, there are first as many time-points as the amount of steps required to check whether 0 is a counterinstance, afterwards there are as many time-points as required to check 1, ... After all these computations have been carried out, I postulate that there is a final time-point where, as usual, the reader exists and waits for a signal from the

machine. In world $w$, the reader can then determine the Conjecture's truth-value by the given signal's presence or absence. By construction, the machine will not find a counterinstance and, as a consequence, it will not send a signal.

Observe that, perhaps, for some $m \in M$, the machine might have to execute a non-standard amount of steps in its computation history before halting. In turn, this corresponds to a non-standard amount of time-points so that in $w$ time might contain a non-standard order. Still, this is no obstacle: the Argument from Metaphysically Possible Supertask Computations teaches us that all that matters for determinacy is that $w$ be just a *metaphysically* possible world as opposed to a physically possible world, meaning that we are free to postulate all kinds of bizarre orders for time.

By appealing to the Argument from Metaphysically Possible Supertask Computations and the metaphysically possible computation carried out in $w$, we can infer the determinacy of Goldbach's Conjecture. It is in this sense that I claim that the relevant programme in Benacerraf and Putnam's argument does not need to be running inside the standard model: computability assertions satisfied in non-standard models give rise to metaphysically possible decision procedures. We can then replace $\Diamond_M$ Backwards Computation with one of these possibilities in the Argument from Metaphysically Possible Supertask Computations, and in this way secure semantic determinacy. I think it is this exact feature of their argument which exposes Benacerraf and Putnam to refutation, as I explain next.

### 3.5.2 A New Objection to Benacerraf and Putnam's Argument

Suppose, perhaps *per impossibile*, that $\forall x \mathsf{GB}(x)$ is independent of $\mathsf{PA}$ and, therefore, that $\mathbb{N} \models \forall x \mathsf{GB}(x)$. We can take note of some simple model-theoretic facts that now follow. First, $\mathbb{N} \models \mathsf{GB}(\overline{n})$ for every $n \in \mathbb{N}$. By the Completeness of first-order logic and the Downward Löwenheim-Skolem Theorem, there is a non-standard countable model $\mathfrak{M} \models \mathsf{PA} + \neg\forall x \mathsf{GB}(x)$. Recall that $\Delta_0$-truth is absolute, that is: for any $\varphi \in \Delta_0$ and for any model of arithmetic $\mathfrak{A}$ it is the case that $\mathbb{N} \models \varphi \Rightarrow \mathfrak{A} \models \varphi$. We then have that $\mathfrak{M} \models \mathsf{GB}(\overline{n})$ for every $n \in \mathbb{N}$. Hence, there is some non-standard $m \in M$ which is in $\mathfrak{M}$ a witness for $\neg\forall x \mathsf{GB}(x)$.

The idea now is to run the programme searching for a counterinstance to Goldbach's Conjecture inside both $\mathbb{N}$ and $\mathfrak{M}$, respectively. We can then consider metaphysically possible worlds where these computations are carried out. Finally, we apply the Argument from Metaphysically Possible Supertask Computations to secure semantic determinacy. To this end, consider first the programme running inside $\mathbb{N}$. There is a metaphysically possible world $w_{\mathbb{N}}$ where the computation is carried out. Since, by assumption, $\mathbb{N} \models \forall x \mathsf{GB}(x)$, in $w_{\mathbb{N}}$ the computation does not find a relevant counterinstance. As per Benacerraf and Putnam's argument, we should take this as a fact determining that there is no counterinstance to the Conjecture. Therefore, from the programme running inside $\mathbb{N}$, we infer:

In $w_{\mathbb{N}}$, for any intended model $\mathfrak{M}$ and domain $M$ it follows:
For every $m \in M : \mathfrak{M} \models \mathsf{GB}(x/m)$ (Determinacy in $w_{\mathbb{N}}$)

By Generalization we conclude that Goldbach's Conjecture is determinately true in that world:

In $w_{\mathbb{N}}$, for any intended model $\mathfrak{M} \models \forall x \mathsf{GB}(x)$

Since arithmetical relations do not change across worlds, we get that Goldbach's Conjecture is determinately true *simpliciter*. This is just the reasoning behind the Argument from Metaphysically Possible Supertask Computations.

Consider now the programme running inside $\mathfrak{M}$. Similar to before, there is a metaphysically possible world $w_{\mathfrak{M}}$ where the computation is carried out. Since, by assumption, $\mathfrak{M} \models \mathsf{GB}(\overline{n})$ for every $n$, it follows that there is no standard counterinstance to the Conjecture. Still, by construction, there is a non-standard $m \in M$ which is a witness to $\neg \forall x \mathsf{GB}(x)$. This means that in $w_{\mathfrak{M}}$ the computation finds a counterinstance in the non-standard part. And again, as per Benacerraf and Putnam's argument, we should take this as a fact determining that there is

some counterinstance to the Conjecture. From the programme running inside $\mathfrak{M}$, we infer:

In $w_{\mathfrak{M}}$, for any intended model $\mathfrak{M}$ and domain $M$ it follows:
For some $m \in M : \mathfrak{M} \models \neg \mathsf{GB}(\mathsf{x}/m)$ (Determinacy in $w_{\mathfrak{M}}$)

By Generalization we conclude that Goldbach's Conjecture is determinately false:

In $w_{\mathfrak{M}}$, for any intended model $\mathfrak{M} \not\models \forall \mathsf{x} \mathsf{GB}(\mathsf{x})$

Equivalently, in $w_{\mathfrak{M}}$, for any intended model $\mathfrak{M} \models \exists \mathsf{x} \neg \mathsf{GB}(\mathsf{x})$

Since arithmetical relations do not change across worlds, we get that Goldbach's Conjecture is determinately false *simpliciter*. Finally, joining the two results yields a case of *overdetermination*: not only do we secure the determinacy of Goldbach's Conjecture alright, but together the different computations prove the Conjecture to be both determinately true and determinately false. But this is an impossible result. After all, if a sentence is overdetermined then there are (intended) models of arithmetic both satisfying and falsifying $\forall \mathsf{x} \mathsf{GB}(\mathsf{x})$. Benacerraf and Putnam's argument produces an absurd case of overdetermination!

I think that the conclusion is now clear. I take the above result as *reductio* of the idea that failure of a computation in finding a counterinstance entails Determinacy in $w_{\mathbb{N}}$ — just because the computation never finds a counterinstance this does not mean that it is *determinate* that there is no counterinstance.[12] At this point, recall premise (iii) of the Simple Argument:

| The Simple Argument from Supertask Computations |
| --- |
| (iii) If the computation never finds a counterinstance then there is a (conjunctive) fact of the matter — to wit, the fact that the machine never finds a counterinstance, together with the relevant metasemantic facts about how certain facts determine a sentence's truth-value — *determin-* |

[12] Strictly speaking, it can also be taken as a reductio of the analogous idea that whenever a computation does find a counterinstance this entails Determinacy in $w_{\mathfrak{M}}$. But it seems that those two entailments stand or fall together.

> *ing* the Conjecture's truth-value. That is, we should take the fact that the computation never finds a counterinstance as a constraint on our arithmetical practice so that in any intended model each relevant instance of $\mathsf{GB}(x)$ is true and, as a result, that $\forall x \mathsf{GB}(x)$ is true.

According to premise (iii), if the computation never finds a counterinstance this is evidence for the determinacy of each relevant instance. But this is precisely the entailment we have just rejected. This means that the Simple Argument from Supertasks Computations, together with the Argument from Physically Possible Supertask Computations and the Argument from Metaphysically Possible Supertask Computations which contain it, are all unsound: computations are in general silent on questions of determinacy thus refuting Benacerraf and Putnam's argument. In the remainder, I discuss possible replies to this objection.

### 3.5.3 Replies

#### Reply 1: Physical Possibility

My objection makes essential use of the metaphysical possibility of the relevant computations. One reply now has it that what matters for determinacy is the output of computations which are physically possible, as opposed to merely metaphysically possible. Indeed, it might be thought that the concept of a computable procedure presupposes the procedure's physical possibility. It might be even suggested that this is implicit in Turing's (1937) famous analysis of a computation as a task capable of being carried out by an idealised human clerk working mechanically and without creativity.

On the other hand, in regards to the computations defined relative to non-standard models, the computation history of the programme on a given input might be coded by a non-standard element. For that input, the computation will carry out a non-standard amount of steps before halting. In turn, this means that time will have a non-standard order in the corresponding metaphysically possible world where the computation is performed. If we assume that a world where

time contains a non-standard order is not physically possible, then requiring that only the output of physically possible computations be relevant for truth determinacy will have, in effect, the result of dismissing the metaphysically possible but physically impossible computations as irrelevant to questions about determinacy. The computation which finds a counterinstance to Goldbach's Conjecture after a non-standard amount of steps is all well and good, but it is by assumption physically impossible and so extraneous to the question of whether Goldbach's Conjecture is determinate or indeterminate. If so, my objection to Benacerraf and Putnam's argument errs in inferring the indeterminacy of Goldbach's Conjecture from physically impossible computations.

I find three three general difficulties with this reply. First, it is unclear why a non-standard order for time should be physically impossible. Second, the reply makes the Argument from Metaphysically Possible Supertask Computations collapse into the Argument from Physically Possible Supertask Computations, sacrificing the advantage in taking the modality to be of a broader kind. That is, Benacerraf and Putnam's argument will now carry commitments regarding which computations and which spacetime configurations are physically possible. In this way, Benacerraf and Putnam (together with my hypothetical objector) will be forced to reintroduce assumptions about the physics of spacetime which the Argument from Metaphysically Possible Supertask Computations was designed to avoid.

Finally, the reply seems to get the order of explanation backwards: if we already harbor doubts about our arithmetical concepts not being sufficiently determinate, are we to suppose that our concepts about what counts as physically possible are any better? Especially when physical theories presuppose much more than elementary number theory? This last objection might sound familiar. I noted in Chapter 1 section 1.2.3 that, according to Field ([1998] 2001b), semantic determinacy is a plausible thesis. He grounds this claim on the fact that a theory of arithmetic extended with a 'finiteness-operator' can yield categoricity (e.g. the theory of PA + 'Every number is finitely-far from 0'). In order to determine the extension of the finiteness-operator, Field introduces his so-called 'cosmological assumptions' about the structure of time. Very loosely, the idea is the following. Let $\varphi(X)$ abbreviate '$X$ is a set of events such that (i) $X$ has an earliest

and a latest member, and (ii) any two members of $X$ occur at least one second apart'. If we now assume that there is no finite bound on the size of the sets satisfying $\varphi(X)$, and that any set satisfying $X$ will have only finitely many members (i.e. time is Archimedean) it will be possible to define a new finiteness-operator $F(X)$ such that: for a set $S$, $F(S)$ if and only if (i) there is some $Z$ such that $\varphi(Z)$ and (ii) there is an injectible function mapping $S$ into $Z$. Field crucially assumes that the extension of much of our physical vocabulary is determinate. Given how finiteness is construed, it seems that any indeterminacy in this concept will be traceable to a deviant interpretation of $\varphi(X)$ and of the physical terms that appear in the latter (such as 'event'). Fortunately, by assumption, we can dismiss these deviant interpretations of the physical vocabulary so that finiteness will be determinate too. Still, if it is already problematic to suppose that our practice fixes the interpretation of 'finiteness', it seems that a similar assumption about our physical vocabulary should be equally questionable. We can find this objection in Parsons who writes:

> Concerning a nonstandard model of S, Field claims that a set of events whose "finiteness" is witnessed by a nonstandard integer will either have to contain objects satisfying 'events' that are not events, or pairs $e$, $e'$ of events satisfying 'earlier than' or 'at least one second apart' such that $e$ is not earlier than $e'$, or $e$ and $e'$ are not one second apart. Consider an 'event' in the sense of the model that unfolds at a stretch of time in the nonstandard part. Field's assumption is that that is not an event. I find it hard to see how someone could accept that assumption who does not already accept some hypothesis that rules out nonstandard models as unintended on mathematical grounds. If our powers of mathematical concept formation are not sufficient to do the latter, then why should our powers of physical concept formation do any better? We are not talking about events in a commonsense context, but, rather, in the context of a physical theory developed in tandem with sophisticated mathematics' (Parsons, 2007, 290).

*Ditto* in our case.

**Reply 2: Independence**

I have assumed that the sentence $\forall x\mathsf{GB}(x)$ is independent of PA. Since the assumption is purely speculative, my objection is itself conditional on a purely speculative assumption. Worse, because I had assumed that arithmetical relations hold of necessity, it follows that if $\forall x\mathsf{GB}(x)$ is provable in PA then it is necessarily so. Therefore, a counterfactual whose antecedent is the independence of $\forall x\mathsf{GB}(x)$ might be a counterpossible. And under the familiar analysis of counterfactuals, any subjunctive conditional with an impossible antecedent is vacuously true. This means that *perhaps* any claim follows from the supposition that $\forall x\mathsf{GB}(x)$ is independent so that my objection ends up not proving anything, really.

This worry is quickly evaded. It suffices to consider a sentence of whose independence we are more certain. To this end, consider $\mathsf{Con(PA)} := \neg\exists x\mathsf{Proof}(x, \overline{\ulcorner 0{=}1 \urcorner})$. On the assumption that PA is $\omega$-consistent, by Gödel's Second Incompleteness Theorem we have that $\mathsf{PA} \nvdash \mathsf{Con(PA)}$ and $\mathsf{PA} \nvdash \neg\mathsf{Con(PA)}$. Since the formula $\mathsf{Proof}(x,y)$ represents the proof relation we also have that $\mathsf{PA} \vdash \neg\mathsf{Proof}\ (\overline{n}, \overline{\ulcorner 0{=}1 \urcorner})$ for every $n$ and, therefore, that $\mathbb{N} \models \mathsf{Con(PA)}$. On the other hand, by the Completeness of first-order logic and the Downward Löwenheim-Skolem Theorem, there is a non-standard countable model $\mathfrak{M} \models \neg\mathsf{Con(PA)}$. We also know that $\mathfrak{M} \models \neg\mathsf{Proof}\ (\overline{n}, \overline{\ulcorner 0{=}1 \urcorner})$ for every $n$. It follows that there is a non-standard $m \in M$ which in $\mathfrak{M}$ is a witness for $\exists x\mathsf{Proof}(x, \overline{\ulcorner 0{=}1 \urcorner})$.

We now reason exactly as before. Let $e$ be a programme searching for a proof of $0{=}1$. If $e$ runs inside $\mathbb{N}$ then it never finds a relevant proof. By Benacerraf and Putnam's argument, it follows that $\mathsf{Con(PA)}$ is determinately true. But why does the computation never find a proof? This is explained by the fact that there is no standard number which codes a proof of $0{=}1$. After all, the sentence is constructed in such a way that there can be no relevant PA-proof on pain of PA's own inconsistency. But this only passes the buck: why is the computation looking only at standard numbers or, more precisely, why is $e$ running inside the standard model in the first place? There is nothing in Benacerraf and Putnam's argument forbidding $e$ to run inside the model $\mathfrak{M} \models \mathsf{PA} + \neg\mathsf{Con(PA)}$ instead. In this new case, if it 'waits' a non-standard amount of time, the programme $e$ eventually finds a non-standard proof of $0{=}1$. And if, again as per Be-

nacerraf and Putnam's argument, the result of a computation is enough to secure determinacy, it follows that $\mathsf{Con(PA)}$ is determinately false. The sentence $\mathsf{Con(PA)}$ is yet another example of a determinately true and determinately false sentence.

### Reply 3: Non-Standard Proofs

My answer to reply 2 relied on a model 'thinking' that it is inconsistent. If $\mathfrak{M} \models \neg\mathsf{Con(PA)}$ then there is an element in the domain which codes a proof (in the model's sense) of an inconsistency. However, if PA is consistent then there can be no proof of $0{=}1$. This means that in $\mathfrak{M}$ the witness for $\exists x\mathsf{Proof}(x, \overline{\ulcorner 0{=}1 \urcorner})$ is a non-standard element $m \in M$ coding a proof of non-standard length. The issue now is that proofs are supposed to be finite strings of symbols, so that it seems misguided to say that whenever a programme searching in $\mathfrak{M}$ finds a non-standard (code of a) proof of $0{=}1$, the programme did its job and found a 'real' proof of $0{=}1$. On the contrary, whenever the programme finds a non-standard proof of $0{=}1$, it seems that *that* proof is not a real counterinstance to $\neg\exists x\mathsf{Proof}(x, \overline{\ulcorner 0{=}1 \urcorner})$. After all, a counterinstance to $\neg\exists x\mathsf{Proof}(x, \overline{\ulcorner 0{=}1 \urcorner})$ is a proof of $0{=}1$ whereas the non-standard number found by the programme is not a real proof. Hence, the result of the computation is not evidence that the sentence $\neg\exists x\mathsf{Proof}(x, \overline{\ulcorner 0{=}1 \urcorner})$ is determinately false.

Clearly, this reply is conditional on the idea that a proof coded by a non-standard number is not a 'real' proof. But it goes without saying that the concepts of a 'real proof' as well as of a 'real counterinstance' are painted with a very broad brush. And so, absent further elaboration, it is difficult to understand the reply's 'real' significance! The only way I manage to make sense of these concepts is by reference to arithmetical facts of the matter, as discussed in section 3.2. The idea is the following: suppose a model thinks that some element codes a proof of $0{=}1$; if the proof is a real proof then there is an arithmetical fact of the matter, namely, the fact that there is a proof of $0{=}1$; if the proof is not a real proof then there is no such fact. To repeat, if it is a real proof then there is a correspondence between the arithmetical facts and what the model thinks about provability. Granted still pretty vague, I think this picture pins down the right sort of intuition: non-real proofs, non-standard proofs, are mere logical oddities which

do not capture the relevant arithmetical facts.

Recall that intended models are, loosely, the models that capture the relevant arithmetical facts. Arguably, on the assumption that any non-standard proof is not a real proof, it seems that any model containing a non-standard proof — or, more precisely, containing a non-standard element which is a proof in the model's sense — captures the wrong facts about provability and, therefore, is unintended. Now, the Overspill Theorem guarantees that in a non-standard model a formula $\varphi(x)$ is satisfied by a non-standard number whenever it is satisfied by arbitrarily large standard numbers. And because the predicate $\mathsf{Proof}(\mathsf{x}, \overline{\ulcorner 0{=}0 \urcorner})$ is satisfied by arbitrarily large standard numbers the Overspill Theorem proves the existence of a non-standard proof of $0{=}0$ in *every* non-standard model. Every non-standard model becomes unintended which, in effect, isolates the standard model up to isomorphism as the sole intended model. Similar to the argument from the $\omega$-rule and TA, this requirement will characterize the intended models up to elementary equivalence — all intended models will have the same first-order theory — securing semantic determinacy independently of the output of the relevant computations. This again renders Benacerraf and Putnam's argument redundant.

### Reply 4: Tennenbaum's Theorem

I now come to the last and as I see it most natural reply. When describing the computation which finds a counterinstance to $\mathsf{Con}(\mathsf{PA})$ (or Goldbach's Conjecture) the background model was non-standard: the programme *e* computed on numbers and relations interpreted in a non-standard way. However, since any non-standard model codes non-computable information it follows that even the interpretation assigned to the basic operations of addition and multiplication is non-computable — that is, the model is non-computable. But the process by which the machine finds a (code of a) proof of $0{=}1$ should be presumably a computable procedure. And as a computable procedure, it should be reproducible by an idealized calculator working in a mechanical fashion, again as per Turing's (1937) analysis. Since the interpretation assigned to addition and multiplication in a non-standard model are non-computable, this casts doubt on the idea that the computation which finds a counterinstance to $\mathsf{Con}(\mathsf{PA})$ is reproducible

by a calculator and therefore that it is 'really' computable. Hence, my hypothetical objector would continue, we can dismiss whatever the programme does inside the non-standard model as irrelevant to questions about determinacy. More precisely, only computations running inside computable models are relevant to semantic determinacy.

Much like reply 3, this new objection appeals to a fuzzy divide between real and not real computations. I am happy to brush this aside for I think there are deeper issues. The reply asserts that non-standard models are non-computable. But why this assumption in the first place? I think that Benacerraf and Putnam (together with my hypothetical objector) have an easy answer. They can say: 'Non-standard models are non-computable by pure mathematical fact. It even has a name, look! It's just Tennenbaum's Theorem.'

**Theorem 3.5.1** (Tennenbaum's Theorem)**.** *If* $\mathfrak{M} = \langle \mathbb{N}, 0^{\mathfrak{M}}, S^{\mathfrak{M}}, +^{\mathfrak{M}}, \times^{\mathfrak{M}} \rangle$ *is a non-standard model of PA then addition* $+^{\mathfrak{M}}$ *is not computable. Equivalently, if* $\mathfrak{M}$ *is a model of PA and* $+^{\mathfrak{M}}$ *is computable then* $\mathfrak{M}$ *is isomorphic to the standard model* $\mathbb{N}$*. (For a proof see Kaye, 2011)*

The theorem singles out some class of models isomorphic to $\mathbb{N}$ as the computable models. Now, observe that the present objection argues that only computations running inside computable models are relevant to questions of determinacy. But once this requirement is in place, it is very difficult to see what stops us from taking the extra step of demanding altogether that only computable models are intended. After all, only programmes running inside computable models are 'really' computable, meaning that if a non-computable model thinks that there is a halting computation history for a programme on a given input (as in $0$—$\mathfrak{M}$) this does not say anything about 'real' computability — it does not capture the relevant facts about computability. Hence, it seems we should dismiss non-computable models when it comes to our judgments about determinacy. (At this point we might worry that 'real computability' is too metaphorical or imprecise of a concept to do any serious work. I would agree and add that since the concept is essential for the opponent to elaborate her reply, if it is abandoned then the reply is abandoned too.) However, the claim that only computable models are intended entails the extra claim that only models isomorphic to the standard model are intended. The entailment is established precisely by Tennenbaum's Theorem. As a result, this way of restricting the intended models will

again characterize them up to elementary equivalence and secure semantic determinacy. And this situation just repeats a dialectic which is by now familiar: the success of Benacerraf and Putnam's argument hinges on extra theoretical resources — in this case, the assumption that only computations running inside computable models are relevant to determinacy — which, if introduced, already secures determinacy single-handedly, rendering their argument redundant. (This response is inspired by the so-called Argument from Tennenbaum's Theorem. See Chapter 2 and the references therein for elaboration.)

Resorting to Tennenbaum's Theorem invites still a second problem. I argued in Chapter 2 that the mathematical structuralist cannot avail herself of Tennenbaum's Theorem when trying to explain which models are intended for arithmetic on pain of redundancy. And I think that the argument I gave there extends to our present case. To explain. We saw in that chapter how we needed to distinguish between two senses of computability: a function/set is externally computable if it is computed by a Turing machine programme inside the standard model; and a function/set is internally computable if it is computed by a Turing machine programme inside a non-standard model. And we also saw how Tennenbaum's Theorem assumes a privileged concept of computability: a function is computable-in-the-Theorem's-sense if and only if it is computable in the external sense.

Now, recall that my hypothetical objector wanted to say that only decision procedures running inside computable models matter for semantic determinacy. This, she added, is justified by the fact that the interpretation given to basic arithmetical operations by a non-standard model is non-computable as given by Tennenbaum's Theorem. But it is now clear that the Theorem itself, and so the opponent too, are privileging the concept of external computability. Yet, saying that assertions of computability or non-computability for any function/set are to be judged solely relative to how a particular machine performs inside the standard model comes pretty close to saying that an halting computation history inside a non-standard model, and possibly involving a non-standard number of steps, is not a 'real' computation. But this last claim is precisely what my hypothetical objector has to argue for! Simply put, it is circular to use Tennenbaum's Theorem in this case.

## 3.6 Conclusion

According to the semantic determinacy thesis, there are facts determining any arithmetical sentence's truth-value or, equivalently, any sentence is determinately true or determinately false. The thesis dovetails nicely with our intuitions about arithmetical truth, making it an attractive position in the philosophy of mathematics. Benacerraf and Putnam ([1964] 1983) give an argument for semantic determinacy grounded on computations akin to supertasks. In this chapter I have argued against their proposal. Sections 3.1 and 3.2 were introductory. In section 3.3, I explored several formalisations of Benacerraf and Putnam's original argument. The section culminated with the formulation of the *Argument from Metaphysically Possible Supertask Computations* which dispensed with supertasks as well as any strong assumptions about the physics of spacetime. For these reasons, it is hoped that subsequent discussion on Benacerraf and Putnam's argument takes this stronger version as a starting point. In section 3.4, I addressed Warren and Waxman's (2020b) objection from the $\omega$-rule, and argued that they fail to take into account Benacerraf and Putnam's idea that if the computation never finds a counterinstance then this should be taken as evidence that there is no counterinstance in any intended model — as opposed to just no *standard* counterinstance in any intended model. Finally, in section 3.5, I introduced a new objection, to wit: Benacerraf and Putnam's argument extends to computations defined relative to non-standard models, producing bizarre cases of overdeterminacy. I discussed possible replies and found them all wanting.

One way of reading my new objection is as showing that the output of a computation is not sufficient to determine truth and, *a fortiori*, that Benacerraf and Putnam's argument is not sufficient to secure semantic determinacy. By itself, this is not a complete refutation of their argument. This is because the argument might be strengthened with an extra condition or requirement which closes this explanatory gap, permitting the passage from computability to semantic determinacy. Indeed, excepting response 2, this strategy was followed in all of the earlier replies. Now, even though I cannot anticipate every possible counterargument, I think that no strengthening of Benacerraf and Putnam's argument will be successful in answering my objection. To

see why, let us go back to replies 3 and 4: there, the idea was that Benacerraf and Putnam's argument should be reformulated as saying that only the output of a computation defined relative to models that do not contain non-standard proofs (reply 3) or that are computable (reply 4) is sufficient to determine truth. However, I also explained how the introduction of these extra requirements grants us the theoretical resources needed to characterise intended models up to elementary equivalence. More generally, it seems that any extra condition or requirement restricting the class of computations to which Benacerraf and Putnam's argument applies (e.g. a requirement isolating the class of relevant computations as those running inside computable models) is justified only on the basis of a similar requirement that characterizes intended models up to elementary equivalence (e.g. a requirement to the effect that every intended model be computable). Hence, I conjecture that my objection is only avoided after the introduction of explanatory resources which render Benacerraf and Putnam's original argument redundant. I conclude that there is no semantic determinacy from supertask computations.

# Chapter 4

# Krajewski's Theorem, Arithmetical Truth and Paraconsistency

**Abstract**

Satisfaction classes provide a formal characterisation of the concept of truth. Krajewski's Theorem teaches us that there are models of arithmetic admitting mutually inconsistent satisfaction classes. The result can be taken to show that arithmetical ontology is not sufficient to pin down the truth of some formulas. Here, I explore a different reading of the Theorem. I read the phenomenon of mutually inconsistent satisfaction classes as pointing to the existence of formulas which are both true and false. Based on this idea, I propose a paraconsistent semantics for arithmetic and discuss how the new interpretation bears on the claim that arithmetic is determinate. I argue that, in light of Krajewski's Theorem, the concept of determinacy can be understood in terms of the possession of a fixed classical or non-classical truth-value.

## 4.1 Introduction

Arithmetical truth is said to follow from the particular models which ground a sentence's truth, in the following sense: if there is indeterminacy in the truth-value of an arithmetical sentence then there is indeterminacy in the underlying model; equivalently, if the underlying model is

determinate then the truth-value of any arithmetical sentence is determinate as well. This is a natural idea and one which runs as a thread through several debates in the philosophy of mathematics. But it is not the only compelling thesis. Arithmetic is also said to be semantically determinate: every arithmetical sentence is determinately true or determinately false, there being no indeterminate sentences. However popular, the two theses are not beyond doubt.

Satisfaction classes provide a formal characterisation of truth which extends to every sentence in a model's sense. This concept formalizes in a natural way the more familiar metatheoretic concept of satisfaction common to model-theoretic semantics, in the same way that provability predicates formalize the concept of proof. When we ask after the extension of a satisfaction class in non-standard models of arithmetic, Krajewski's Non-Uniqueness Theorem (1974, 1976) teaches us that there are models admitting mutually inconsistent satisfaction classes. From this, it can be argued that mathematical ontology is not sufficient to pin down the truth of some formulas, raising questions about the very determinacy of arithmetical truth.

In this chapter, I explore a different reading of the Theorem and its consequences. I start by explaining how Krajewski's Theorem challenges common theses in the philosophy of mathematics (section 4.2). I then propose to interpret the phenomenon of mutually inconsistent classes as pointing to the existence of formulas which are both true and false. Based on this idea, I introduce a simple paraconsistent semantics for the notion of 'arithmetical truth' similar to the subvaluationist semantics used to model inconsistency and vagueness, as found in Jaśkowski ([1948] 1969) or Hyde (1997), and present some properties of the system (sections 4.3). Afterwards, I address some general objections to my subvaluationist semantics, compare the system with other theories of paraconsistent arithmetic found in the extant literature, and conclude by revisiting the bearings of this interpretation on the claim that arithmetic is determinate (section 4.4). I argue that in light of Krajewski's Theorem the concept of determinacy should be understood in terms of the possession of a fixed classical or non-classical truth-value.

My aim throughout will be purely exploratory: I do not want to take sides in the determinacy/indeterminacy debate nor do I claim that my reading of Krajewski's Theorem is to be preferred, but only that it offers some interesting insights worthy of further discussion.

## 4.2 The Problem

The notion of truth or satisfaction is given by the usual Tarskian model-theoretic account. In this case, truth corresponds to a relation holding between structures and formulas of an appropriate form. This approach is essentially meta-theoretic.

Nonetheless, it is possible to augment PA with a collection of first-order axioms, similar to the ones we use to define satisfaction in the meta-theory, constraining the possible interpretations of a predicate Tr such that the latter becomes an acceptable truth-predicate in the object-language. Once this is done it becomes inevitable to ask after the non-standard sentences falling under the extension of the predicate. Now, that there are in some sense true non-standard sentences is immediate. Peano Arithmetic is already expressive enough to define partial truth-predicates. In particular, there is some $\mathcal{L}_{\mathsf{PA}}$-formula $\mathrm{Tr}_{\Sigma_n}(x)$ representing that '$x$ is the Gödel-code of a true $\Sigma_n$-sentence'. Since the predicate is satisfied by arbitrary large codes of such formulas, an application of Overspill, for instance, tells us that in any non-standard model of PA there is some non-standard number $c$ falling under the predicate. The non-standard model thinks that the number codes a *true* sentence. Implicit in this argument is the idea that the number codes a true *sentence*; that $c$ satisfies some extra predicate, say, $\mathrm{Sent}_{\mathsf{PA}}(x)$ intended to express that 'the sequence with Gödel-code $x$ is a sentence of $\mathcal{L}_{\mathsf{PA}}$'.

Krajewski (1974, 1976) was the first to formulate the concept of 'satisfaction class'. Informally, for a given model, a satisfaction class is a particular subset of the model's domain which can be thought of as the extension of the satisfaction relation for both standard and non-standard sentences. Krajewski went on to prove that there are models of arithmetic with many mutually inconsistent satisfaction classes. This result comes with surprising consequences. Let me explain.

### 4.2.1 The No-Difference Thesis

Truth comes cheap. There are many ways to make first-order Peano Arithmetic true. A classic example comes from Benacerraf (1965) who reminds us that two quite disparate mathematical

ontologies – Zermelo ordinals on one hand, and von Neumann ordinals on the other – induce structures which make the exact same arithmetical sentences true. But there are more general results in the vicinity. The MacDowell-Specker Theorem for instance teaches us that every model of Peano Arithmetic has a proper elementary end-extension (Kaye, 1991, Theorem 8.6). And arithmetic admits models of any infinite cardinality on account of the Löwenheim-Skolem theorems. More simple still, any bijection between some model's domain and another set can be used to induce an isomorphic structure[1]; therefore, any bijection on $\mathbb{N}$ generates many distinct models of $\mathrm{Th}(\mathbb{N})$.[2] What these examples show is that any differences in mathematical ontology, the objects that there are and the way they are related, do not translate in all cases into differences in arithmetical truth.[3] However, the opposite sounds rather bizarre. We expect, after all, for arithmetical truth to be a fixed matter once mathematical ontology is equally fixed. I think this idea is best explained as follows:

> **No-Difference Thesis:** If there is indeterminacy in the truth-value of an arithmetical sentence then there is indeterminacy in the underlying model.

What do I mean by a 'determinate' sentence? Well, a sentence is determinate if and only if there are facts of the matter determining its truth. As stated in Chapter 1, I think this idea is best explained in terms of the extra concept of an 'intended model'. A sentence is determinate

[1] The idea is just to let the interpretation of a non-logical $s$ in the second model $\mathfrak{A}$ be the image of the interpretation in the first model $\mathfrak{M}$ under the bijection $\pi$, so that $s^{\mathfrak{A}} = \pi(s^{\mathfrak{M}})$. Essentially, an interpretation in $\mathfrak{A}$ is defined by 'pushing' the interpretation in $\mathfrak{M}$ through the 'cookie cutter' which is $\pi$. See Button and Walsh (2018, 37) where the result is aptly called the 'push-through construction'.

[2] I follow the innocuous convention of letting $\mathbb{N}$ refer ambiguously to the standard model and its domain.

[3] The stress throughout this paragraph is on *arithmetical* truth. In fact, it is readily seen that changes in ontology easily lead to changes in the truth-value of sentences external to arithmetic. The classic example is again found in Benacerraf (1965): the set-theoretic sentence '$1 \in 3$' will have a different truth-value depending on whether we use the Zermelo or the von Neumann construction of the naturals. This case points to a more general lesson: adding extra expressive power to the language allows for finer distinctions which might not be preserved across different ontologies. At the limit, if our language contains for every possible object $o$ a logical-predicate corresponding to its haecceity – to wit, the property 'being $o$' – then *any* change in ontology will trivially correspond to a change in the distribution of haecceities and so to a change in the truth of sentences predicating these kind of properties. Of course this conclusion hinges on more general debates on the metaphysics of haecceities (see Cowling (2015, 2016) for further discussion and a survey of the literature, and see Shapiro (2006, 133ff) for haecceities within mathematical structuralism). Either way, it is clear why I qualify the claim in the main text to the language of arithmetic.

if and only if all the structures in the intended class of arithmetical models decide it in the exact same way — either they all satisfy or all falsify the sentence. Later we will see how Krajewski's Theorem urges us to revise this notion slightly. For now the following familiar schema suffices:

For $\varphi$ a sentence in the language of arithmetic $\mathcal{L}_{\mathsf{PA}}$:

$$\begin{cases} \varphi \text{ is determinately true if it is satisfied by all intended models} \\ \varphi \text{ is determinately false if it is falsified by all intended models} \\ \varphi \text{ is indeterminate otherwise} \end{cases}$$

Talk of 'an indeterminacy in the underlying model' has to do with the intended models' structure. The underlying model is said to be indeterminate whenever intended models are not pairwise isomorphic. In this sense, the No-Difference thesis states that if there is disagreement among intended models with respect to some arithmetical sentence then intended models are not pairwise isomorphic.

It is worth noting that the thesis comes very close to a supervenience claim. In a way, it states that there is no difference in truth, no disagreement among intended models, without a corresponding difference in those models' structure. Nevertheless, I am wary of presenting it exactly in these terms. And this is because supervenience is often treated as a modal relation. What would a modal version of the No-Difference thesis look like? Perhaps, it would be something like: if in some world $w$ an arithmetical sentence is true while in some other world $v$ it is false, then the intended model grounding the sentence's truth in $w$ is structurally different to the intended model grounding its truth in $v$. The problem, however, is that this seems to make the No-Difference thesis reliant on some ontology of possible worlds. But the fictionalist who does not believe in possible worlds and instead prefixes arithmetical and modal truths with an 'according to the fiction'-operator might want to have a No-Difference thesis of her own, namely: *according to the fiction of mathematics* if there is indeterminacy in the truth-value of an arithmetical sentence then there is indeterminacy in the underlying model. Relatedly, if mathematical truths are necessary, it is equally odd to think that their truth changes across worlds. For these reasons, it is best to understand No-Difference as a conditional as opposed to a modal

statement — still, always bearing in mind that the thesis is akin to a supervenience claim.

### 4.2.2 No-Difference At Work

As I see it, No-Difference plays a significant role in some debates. In the philosophy of mathematics, there is a well-known argument from Gödel's First Incompleteness Theorem to the indeterminacy of the structure of arithmetic. The idea is to start with one of those independent sentences that Gödel, and later Rosser, taught us how to build. Afterwards we observe that the relevant facts of the matter determining arithmetical truth do not fix a truth-value for the independent sentence — arithmetical practice can, the argument goes, be described equally well by two intended models which decide the relevant sentence in opposite ways. It follows that the independent sentence is indeterminate. (This was also how we argued for semantic indeterminacy in subsection 1.2.3.) Finally, we conclude that intended models are not pairwise isomorphic. Operative in the final stage of the argument is the No-Difference thesis: that if there is indeterminacy in the truth-value of some arithmetical sentence then this must be reflected in the indeterminacy of the underlying model. Equivalently, if the underlying model is unique up to isomorphism then No-Difference dictates that arithmetical truth is determinate as well. Curiously, this last inference seems to be the driving intuition behind some of Feferman's views. He writes:

> One might even go farther to say that the picture of sets in the cumulative hierarchy is sufficiently clear that the portion $C_\omega$ of the cumulative hierarchy consisting of the hereditarily finite sets is well determined. Since the natural numbers are extracted from $C_\omega$ in usual developments of set theory, according to this picture every number-theoretical statement provable in ZFC is true. [...] in gross the picture is clear enough to justify confidence in the use of ZFC (and like theories) for deriving number-theoretic results. (Feferman, 1998, 73)

> The conception of the structure N of the natural numbers is not a vague one (at least in my view); statements about N have a definite truth value, and the axioms

> of PA are among those and (on reflection) are evident for it. (Feferman, Friedman, Maddy, & Steel, 2000, 410)

In later writings Feferman is much more direct about the illation:

> The conception of $(\mathbb{N}^+, 1, Sc, <, +, \times)$ is so clear that, again implicitly, at least, there is no question in the minds of (amateur or professional) number theorists – nor should there be any question in our minds – as to the definite meaning of such statements and the assertion that they are true or false, independently of whether we can establish them one way or the other. In other words, in more modern terms, we should accept realism in truth values for statements about this structure [...]' (Feferman, 2011, 14–15)

> In my view, the conception [of the bare structure of the natural numbers] is *completely clear*, and thence *all arithmetical statements are definite*. (Feferman (2013, 7); quoted in Hamkins and Yang (2013, 2))

Similar to Feferman, Martin writes:

> What I am suggesting is that the real reason for confidence in first-order completeness is our confidence in the full determinateness of the concept of the natural numbers. (Martin, 2018, 15)[4]

One way of interpreting the difference between these quotes and the earlier argument from Gödel's First Theorem is that one applies modus tollens whilst the other modus ponens. But common to both is No-Difference.

So much for the philosophy of mathematics. When it comes to philosophy more generally, the thesis looks like an arithmetical variation on a more general idea about truth. Bigelow argues that a change in the truth-value of a formula is always accompanied by a corresponding change in the objects that exist; or, in his famous dictum:

[4]Where '[t]he concept of the natural numbers is *first-order complete*: it determines truth-values for all sentences of the usual first-order language of arithmetic' (2018, 4) and '[...] a structural concept is fully determinate if and only if the concept fully determines a single *isomorphism type*' (2018, 11).

> Truth is supervenient on being. [...] If something is true, then it would not be possible for it to be false unless certain things were to exist which don't, or else certain things had not existed which do. (Bigelow, 1988, 132–133; see also 158–159)

This is a bit strange though. For there are cases where we want to say that truth changes while the objects that exist remain fixed — as when an object simply loses some property or other. To accommodate this intuition Lewis proposes an extension of Bigelow's criterion:

> For any proposition P and any worlds W and V, if P is true in W but not in V, then either something exists in one of the worlds but not in the other, or else some n-tuple of things stands in some fundamental relation in one of the worlds but not in the other. (Lewis, 2001, 612)

In a similar spirit Parsons introduces (though does not explicitly advocate) a 'Truthmaker Principle' which he defines as:

> This assertion, that every true sentence's truth supervenes on the nature of some thing, is what I will mean by the 'truthmaker principle'. To put this another way, for every true sentence, there is some thing such that the sentence cannot become false without a qualitative change, a non-Cambridge change, in that thing. That thing, whatever it is, is the sentence's truthmaker. (Parsons, 1999, 327)

I find Lewis and Parsons' ideas close cousins of the No-Difference thesis: if we take the proposition '1 + 1 = 2' and suppose, perhaps *per impossibile*, that it is true in world W and false in V, Lewis' criterion tells us that there is some object or fundamental relation in one of these worlds not present in the other. In a way, the No-Difference thesis simply interprets this 'some object or relation' in terms of arithmetical models. Similar remarks apply to Parsons' exposition. Nonetheless, the analogy should not be pursued too far. Lewis and Parson's criteria are formulated in a possible world semantics, and we have already seen the difficulties with a world-based

account of No-Difference.[5] My point is simply that both my way of explaining the relation between determinacy of arithmetical truth and the underlying mathematical models, as well as Lewis and Parsons's own way of explaining the relation between truth-more-generally and the features of the world that truth is about see as crucial a relation akin to supervenience. This is a non-trivial feature shared by all our approaches.

Even though this sample is anecdotal I think we can safely assume that something along the lines of our No-Difference thesis is in the arithmetical case widely shared if not only for its sheer intuitiveness. Or, as Hamkins and Yang (2013, 25) write: 'The main concern is whether [...] definiteness-about-truth follows from definiteness-about objects and definiteness-about-structure. *Many mathematicians and philosophers seem to think that it does.*' (my emphasis).

The problem is that Krajewski's Non-uniqueness Theorem seems to go against No-Difference. Since to elaborate on this issue it will be helpful to have a more formal statement of the Theorem at our disposal, I turn to some technical preliminaries in the next subsection.

### 4.2.3 Satisfaction Classes

The base theory, defining the objects over which truth applies, is first-order Peano Arithmetic (PA). The language $\mathcal{L}_{\mathsf{PA}}$ is the language of first-order logic with the usual connectives and extended with the non-logical primitive symbols $\{0, s, +, \times\}$. In PA any computable set of natural numbers is $\Sigma^0_1$-representable. The representability of computable predicates allows us to mimic syntactic properties inside the theory. For a suitable Gödel-coding, let $\mathrm{Sent}_{\mathsf{PA}}(\ulcorner \varphi \urcorner)$ be the arith-

[5] Moreover Lewis (2001, 612) and Parsons (1999, 326) take their criteria to range over only certain kinds of properties: these include such kinds as qualitative, intrinsic, fundamental and natural properties. However, when stating the No-Difference thesis in arithmetic we can ignore such a diverse ontology. Also, under Lewis' criterion the truth-value of a necessary truth trivially supervenes on the existence or properties of some thing. This is because the criterion's antecedent will be vacuously satisfied since the truth-value of a necessary truth does not change across worlds. The problem is that this makes the supervenience of a necessary truth on some underlying ontology just too easy; for a formulation of the problem see Liggins (2005, 109–110). *A fortiori*, if arithmetical truths are necessary, as they are normally taken to be, Lewis' criterion does not give a correct account of supervenience for them. Finally there are many other issues – e.g. negative existentials, the ontology of states of affairs, etc. – related to Lewis and Parsons' criteria which are extraneous to our No-Difference thesis in arithmetic. For further discussion of supervenience relations in the philosophy of truth-makers see, among others, Dodd (2002), Liggins (2005) and Merricks (2007, chapter 4).

metical formula abbreviating '$\varphi$ is a sentence in the language $\mathcal{L}_{\mathsf{PA}}$'. Moreover, for a model of arithmetic $\mathfrak{M} \models \mathsf{PA}$, let $\mathrm{Sent}_{\mathsf{PA}}^{\mathfrak{M}}$ be the set of elements which are sentences in $\mathfrak{M}$'s sense. We can then define:[6]

**Definition 4.2.1** (Full Satisfaction Class)**.** *For a model $\mathfrak{M}$ of PA and a set $S \subseteq Sent_{PA}^{\mathfrak{M}}$, we call $S$ a full satisfaction class for $\mathfrak{M}$ if and only if $S$ satisfies the Tarskian recursive clauses for truth, that is:*

1. *For every $\varphi \in Sent_{PA}^{\mathfrak{M}}$, if $\varphi$ is of the form $s^a(0) + s^b(0) = s^c(0)$ then $\varphi \in S$ if and only if $a + b = c$;*
2. *Similar for other atomic formulas;*
3. *For every $\varphi, \psi \in Sent_{PA}^{\mathfrak{M}}$, the disjunction $\varphi \vee \psi$ is in $S$ if and only if $\varphi \in S$ or $\psi \in S$;*
4. *For every $\varphi, \psi \in Sent_{PA}^{\mathfrak{M}}$, the conjunction $\varphi \wedge \psi$ is in $S$ if and only if $\varphi \in S$ and $\psi \in S$;*
5. *For every $\varphi \in Sent_{PA}^{\mathfrak{M}}$, $\neg\varphi$ is in $S$ if and only if $\varphi \notin S$;*
6. *For every formula $\varphi(v)$ in the sense of $\mathfrak{M}$, the sentence $\exists v \varphi(v)$ is in $S$ if and only if $\varphi(s^m(0)) \in S$ for some $m$ in the domain $M$*

*If there is such a set $S$ we say that $\mathfrak{M}$ admits a full satisfaction class.*

Satisfaction classes are an 'arithmetised' version of truth in the same way that provability predicates are arithmetised versions of proof. Note that, for readability, a lot of coding has been omitted from the definition. In clause (4), the sentence $\varphi \wedge \psi$ is actually a number (a code) in the domain $M$ which is a witness for the relation $\mathrm{Conj}_{\mathsf{PA}}$(x,y,z) — abbreviating 'x is the Gödel-code of a formula in the language $\mathcal{L}_{\mathsf{PA}}$ which is the conjunction of a sentence with Gödel-code y and a sentence with Gödel-code z' — whenever y is $\varphi$ and z is $\psi$. Similar remarks extend to the other clauses. In the remainder, I will identify a sentence with its code.

Now, the first thing to note about these classes is that the meta-theoretic relation of satisfaction and the property of belonging to a satisfaction class coincide on the level of standard sentences. This idea is explained more fully by the next result:

[6] There are several equivalent ways of defining full satisfaction classes. I follow the formulation in Kotlarski (2019, 53–54).

**Lemma 4.2.2.** *Let $\mathfrak{M}$ be a model of PA admitting a full satisfaction class $S$. For a formula $\varphi(v) \in \mathcal{L}_{PA}$ and a standard element $m \in M$ it follows*

$$\mathfrak{M} \models \varphi(\overline{m}) \Leftrightarrow \ulcorner \varphi(\overline{m}) \urcorner \in S$$

*Proof Idea.* If $\varphi$ is atomic the equivalence is established by clauses (1) and (2) in the definition 4.2.1. For sentences of greater complexity the equivalence follows by induction. □

Essentially, for a given model, every satisfaction class agrees with respect to standard sentences. As it will be more clear below, the existence of mutually inconsistent satisfaction classes entails that this agreement does not extend to non-standard sentences. Now define:

**Definition 4.2.3** (Computably Saturated Model)**.** *Let $p(v) := \{\varphi_i(v) : i \in I\}$ be a set of formulas in the language of a model $\mathfrak{M}$ with one free variable $v$ and possibly with a finite set of parameters from $\mathfrak{M}$. We say $p(v)$ is realised in $\mathfrak{M}$ if and only if there is an $m$ in the model's domain such that, for every $i \in I$, $\mathfrak{M} \models \varphi_i(m)$. Moreover we call $p(v)$ a type over $\mathfrak{M}$ if and only if every finite subset of $p(v)$ is realised in $\mathfrak{M}$. And $p(v)$ is a computable type over $\mathfrak{M}$ if and only if $p(v)$ is a type over $\mathfrak{M}$ and $p(v)$ is a computable set. Finally, we call a model $\mathfrak{M}$ computably saturated if and only if it realises every computable type over $\mathfrak{M}$.*

Computable saturation allows for a characterization of models admitting a satisfaction class in terms of both necessary and sufficient conditions. Sufficient conditions are given by the following theorem:

**Theorem 4.2.4** (Kotlarski, Krajewski, and Lachlan (1981); Kaye (1991))**.** *If $\mathfrak{M}$ is a countable computably saturated model of PA then it admits a full satisfaction class $S$.*

Moreover, we have:

**Theorem 4.2.5** (Lachlan (1981))**.** *If $\mathfrak{M}$ is a non-standard model of PA and admits a full satisfaction class then it is computably saturated.*

Together Theorem 4.2.4 and 4.2.5 provide the following equivalence:[7]

**Proposition 4.2.6.** *A countable non-standard model of PA admits a full satisfaction class if and only if it is computably saturated.*

One further result about computably saturated models will prove useful later on.

**Theorem 4.2.7** (Kaye (1991) Theorem 11.4.)**.** *If $\mathfrak{M}$ is a model of PA then there exists a computably saturated elementary extension of $\mathfrak{M}$ of the same cardinality.*

Let us call two full satisfaction classes $S_1$ and $S_2$ over a model $\mathfrak{M}$ *mutually inconsistent* whenever there is some $m \in M$ such that $m \in S_1$ but $m \notin S_2$. If we think of satisfaction classes as taking note of true sentences, the existence of mutually inconsistent satisfaction classes shows that, in a way, there are sentences which are in a sense true but in another sense false. What's more, the sense in which they are true and false is not the sense to which we are accustomed. After all, the Continuum Hypothesis is both true and false in the sense that it is true according to some models of set theory and false according to others. However, this sort of relativisation is not present when talking about mutually inconsistent satisfaction classes. That is, if we think of satisfaction classes as taking note of true sentences, the existence of mutually inconsistent classes shows that, in a way, there are sentences which are both true and false according to the very same model. *Prima facie*, we would think that no model admits mutually inconsistent satisfaction classes. This just makes the next result even more surprising, if not downright bizarre:

**Theorem 4.2.8** (Krajewski (1974, 1976))**.** *If $\mathfrak{M}$ is a model of PA with a full satisfaction class, then there is an elementary extension of $\mathfrak{M}$ with infinitely many full satisfaction classes*

Since theorem 4.2.8 shows that in many cases satisfaction classes are not unique, it is sometimes called Krajewski's Non-Uniqueness Theorem. With a more definite statement of satisfaction classes at hand, we can return to the No-Difference thesis. (In the remainder, I will use 'full satisfaction class' and 'satisfaction class' interchangeably. For a survey of the research in connection to the study of satisfaction classes see Kotlarski (1991, 2019) and Murawski (1997).)

[7] Smith (1987, 1989) built uncountable computably saturated models without a full satisfaction class meaning that the countability assumption in Theorem 4.2.4 cannot be omitted.

### 4.2.4 Krajewski's Theorem and the No-Difference Thesis

If, as we noted, satisfaction classes are to be thought of as representing sets of true sentences then the existence of mutually inconsistent classes for the same model suggests that there is a mismatch between some sentence's truth and the underlying ontology. Let $\mathfrak{M}$ be an intended model of PA and suppose that there are full mutually inconsistent satisfaction classes $S_1$ and $S_2$ — a phenomenon made possible by Krajewski's Theorem. Let $\psi \in \mathrm{Sent}_{\mathsf{PA}}^{\mathfrak{M}}$ be a sentence such that $\psi \in S_1$ but $\psi \notin S_2$. Since both $S_1$ and $S_2$ are satisfaction classes, it seems that we can equally think of $\psi$ as a true sentence in $\mathfrak{M}$, using $S_1$, and as a false sentence, using $S_2$. Different satisfaction classes sort of *push in different directions* leading to indeterminacy over the truth of $\psi$. And since the underlying model of arithmetic is always the same — it is always $\mathfrak{M}$ — the existence of mutually inconsistent satisfaction classes as given by Krajewski's Theorem shows how it is (logically) possible for there to be only one intended model and at the same time indeterminacy about arithmetical truth. Hence, a counterexample to the No-Difference thesis.

It is instructive to note that Hamkins and Yang (2013) have appealed to finer results in connection to Krajewski's to defend a similar conclusion to our own. For a model $\mathfrak{A}$ of ZFC, let $\mathbb{N}^{\mathfrak{A}} = \langle \mathbb{N}, +, \times, 0, 1, < \rangle^{\mathfrak{A}}$ be the canonical model of arithmetic in $\mathfrak{A}$. That is, $\mathbb{N}^{\mathfrak{A}}$ is the structure which the model thinks corresponds to the natural numbers in $\mathfrak{A}$. Hamkins and Yang prove the existence of two ZFC-models $\mathfrak{A}_1$ and $\mathfrak{A}_2$ which agree on the underlying arithmetical ontology and its structure, say, $\mathbb{N}^{\mathfrak{A}_1} = \mathbb{N}^{\mathfrak{A}_2}$, and yet disagree with respect to arithmetical truth. More precisely, they prove that there is some arithmetical sentence $\sigma$ for which $\mathfrak{A}_1 \models$ '$\mathbb{N}^{\mathfrak{A}_1} \models \sigma$' and $\mathfrak{A}_2 \models$ '$\mathbb{N}^{\mathfrak{A}_2} \models \neg\sigma$' (see their Theorem 1). They go on to argue:

> Our main thesis, in contrast, is that the definiteness of the theory of truth for a structure does not follow as a consequence of the definiteness the structure in which that truth resides. Even in the case of arithmetical truth and the standard model of arithmetic $\mathbb{N}$, we claim, it is a philosophical error to deduce that arithmetical truth is definite just on the basis that the natural numbers themselves and the natural number structure $\langle \mathbb{N}, +, \cdot, 0, 1, < \rangle$ is definite. (Hamkins & Yang, 2013,

26)

Even after fixing the canonical model of arithmetic, both models of set theory disagree on the truths about the former structure. To help us wrap our heads around this result, Hamkins and Yang try to provide a more vivid illustration (2013, 26–27). They ask us to imagine that we 'live' inside the world $\mathfrak{A}_1$. This is to be understood, I take it, as the idea that the sets we think there are together with the claims in set theory that we accept as true coincide with the theory of $\mathfrak{A}_1$. Imagine further that we are committed to $\mathbb{N}^{\mathfrak{A}_1}$ as the unique, or definite, structure of the natural numbers. Their theorem now teaches us that it is illicit to move from this assumption to the further idea that in some other world $\mathfrak{A}_2$ with the same arithmetical structure, arithmetical truths are decided in the same way: 'It seems that, living in $M_1$ [$\mathfrak{A}_1$ in our terminology], we could have a clear and distinct concept of the natural numbers, yet meanwhile, there is another world out there, $M_2$ [$\mathfrak{A}_2$], with the same natural numbers as we have and yet which has different arithmetical truths' (2013, 27).

Regardless of what Hamkins and Yang's arguments do or do not show, it is clear that our own counterexample to No-Difference is not quite right — although I think that the problem is mostly one of terminology. By Lemma 4.2.2, $\psi$ has to be a non-standard sentence. And as we have defined it, determinacy is about there always being a fact of the matter determining the truth-value of a sentence in the language of arithmetic $\mathcal{L}_{\mathsf{PA}}$. Since $\psi$ is not a well-formed-formula in $\mathcal{L}_{\mathsf{PA}}$, we have not really shown that $\psi$ is indeterminate. Determinacy as defined does not actually apply to $\psi$.

Nonetheless, doesn't this simply reveal that something is amiss with the concept of determinacy as defined? And that the concept should be extended to include non-standard sentences like $\psi$? After all, absent some reason not to take $\psi$ into account in our discussion of determinacy, the claim that arithmetic is determinate is as impressive as saying that the Continuum Hypothesis (CH) is determinately true while adding in the same breath that the concept of set-theoretic determinacy does not take into account what follows from models of ZF+¬CH!

But perhaps there is a reason for excluding sentences like $\psi$ when defining determinacy. Perhaps, that is, we can say that determinacy is about there being facts of the matter about whether

*sentences* have fixed truth-values and, moreover, that $\psi$ is not a *real* sentence. For instance, in the spirit of Krajewski's Theorem, Button and Walsh (2018) explain how an extension of PA with the local disquotation schema – that is, a collection of axioms of the form $\mathrm{Tr}(\ulcorner \varphi \urcorner) \leftrightarrow \varphi$ for $\varphi$ a sentence in $\mathcal{L}_{\mathsf{PA}}$ and Tr a predicate – allows for models admitting mutually inconsistent interpretations of the new predicate. Commenting on this result, they write:

> [...] we can *either* regard $a$ as true, or regard $a$ as false, depending on whether we interpret the satisfaction class, $Tr$, via $V_1$ or $V_2$. Otherwise put: semantic properties seem to float completely free from any other properties. This is a little mind-boggling. (Button & Walsh, 2018, 219)

Afterwards, they prove a result similar to Lemma 4.2.2 and continue:

> In this light, indeed, Corollary 9.4 [which proves the existence of mutually inconsistent interpretations of the new predicate] is not even surprising. We have repeatedly seen that first-order theories have 'pathological' models. Corollary 9.4 just presents us with a very particular 'pathological' model of a formal theory of truth. Moreover, its particular 'pathology' can be traced specifically to the fact that it contains an a arithmetically non-standard entity. (*Ibidem*)

As I read them, Button and Walsh start by noting that the existence of mutually inconsistent interpretations of the new predicate seems to undermine some principle akin to the No-Difference thesis — 'semantic properties seem to float completely free from any other properties'. They add that the problematic sentence is actually an 'arithmetically non-standard entity' and, therefore, not a real sentence so that the putative counterexample can be dismissed — 'it is not even surprising'. Indeed, commenting on Hamkins and Yang's (2013) argument they also write:

> However, we want to take issue with Hamkins and Yang's informal claim that $\sigma$ is an *arithmetical sentence*. This is quite misleading. [...] $\sigma$ must be a *non-standard*

> element [...] once the result has been stated in this way, it provides no reason whatsoever to worry that we might somehow agree on the natural numbers but disagree on which arithmetical *sentences* are true. [...] Insofar as we understand the result — i.e. insofar as we speak the language in which the result is stated and proved — we can see that it says *nothing about sentences*. (Button & Walsh, 2018, 221; their emphasis)

Button and Walsh's own reasons for dismissing non-standard sentences are complex, going much beyond the scope of this paper (see their 2018, 210ff for discussion). But observe that if the distinction between real and non-real sentences can be drawn convincingly, it becomes easy to argue for arithmetical determinacy. This is because, again by Overspill, any non-standard model thinks that there is a non-standard sentence. In this case, any non-standard model is wrong about which sequences of symbols are in point of fact (real) arithmetical sentences. These models can then be dismissed, making the standard model up to isomorphism the sole intended. From this, determinacy as defined quickly ensues.[8] In this sense, taking a side in the question about real/non-real sentences is already taking a side in the determinacy/indeterminacy debate. Seeing that my aim in this chapter is purely exploratory, I will assume that Button and Walsh's reasoning can be dismissed and that $\psi$ is a real sentence; or, more precisely, that a satisfactory explanation of determinacy should take $\psi$ into account.

Given this, how should we include $\psi$ and other non-standard sentences in a more refined account of determinacy? Well, I had written that indeterminacy is due to disagreement among intended models: a sentence is indeterminate *if and only if* it is satisfied by some intended models and falsified by others. The biconditional will have to be dropped to a simple conditional; in particular, we should get rid of the left-to-right direction entirely. The reason is straightfor-

[8] This was also one of the arguments presented in Chapter 3 subsection 3.5.3. There I noticed that if we think that proofs coded by a model's non-standard elements are 'not real' proofs then only the standard model up to isomorphism is intended. This is because the Overspill Theorem guarantees that in a non-standard model a formula $\varphi(x)$ is satisfied by a non-standard number whenever it is satisfied by arbitrarily large standard numbers. Since the predicate $\mathsf{Proof}(\mathsf{x}, \ulcorner \overline{0=0} \urcorner)$ is satisfied by arbitrarily large standard numbers the Overspill Theorem proves the existence of a non-standard proof of $0=0$ in every non-standard model of arithmetic, which arguably renders every non-standard model unintended.

ward. Non-standard sentences cannot flank the $\models$–sign, so that a definition of determinacy solely about what models do satisfy/falsify does not leave space for non-standard sentences to play any meaningful role — with the biconditional in place it would not even make sense to say that $\psi$ is determinate or indeterminate. On the contrary, I propose we drop the left-to-right direction and add the following requirement:

> **Non-Standard Indeterminacy Definition:** For $\mathfrak{M}$ an intended model and $\varphi \in Sent_{\mathsf{PA}}^{\mathfrak{M}}$ a (possibly non-standard) sentence, if $\mathfrak{M}$ admits full satisfaction classes $S_1$ and $S_2$ such that $\varphi \in S_1$ and $\varphi \notin S_2$ then $\varphi$ is indeterminate for $\mathfrak{M}$.

The additional definition yields a new sufficient condition for a sentence being indeterminate, taking into account the idea that both non-standard sentences and satisfaction classes should be taken seriously when evaluating our claims to determinacy. Moreover, the new condition is consistent with there being sentences whose indeterminacy is explained in the old-fashioned way in terms of what models do satisfy/falsify. Finally, with the improved account of arithmetical determinacy at hand, we can see how $\mathfrak{M}$ provides a counterexample to the No-Difference thesis. Even after assuming that $\mathfrak{M}$ is the sole intended model, $\psi$ is indeterminate as per the Non-Standard Indeterminacy Definition. But there is no comparable indeterminacy when it comes to the underlying model: the unique model $\mathfrak{M}$. And, of course, this conclusion is not limited to a peculiarly shady sentence $\psi$. The result exacerbates quite quickly on account of the infinitely many satisfaction classes available.

What's more, by construction, a counterexample to No-Difference is a counterexample to the idea that arithmetical truth is determinate. It is curious that the way in which Krajewski's Theorem challenges arithmetical determinacy is so unlike other arguments against determinacy found in the literature. As we saw in Chapter 1, sometimes failure of determinacy is due to a lurking plurality of arithmetical universes. According to plenitudinous platonism, for each model of arithmetic, there is a cluster such that the relations obtaining among the objects in the cluster correspond to the relations obtaining among the elements in the model's domain. In that case, independent sentences come out as indeterminate — the facts about arithmetical reality are too numerous to ever fix a unique truth-value for sentences of this kind. Krajewski's

Theorem opens up a new kind of indeterminacy diverging from the plenitudinous platonist. In this new case, even after assuming the existence of a unique model, in the sense of having a fixed ontology of mathematical objects and fixed arithmetical relations holding over them, there might still be sentences whose truth is not decided in any particular way. This is because those sentences can be regarded as true in virtue of belonging to some satisfaction class, and at the same time regarded as false in virtue of failing to belong to some other class. There does not seem to be any extra fact of the matter that we can appeal to in order to determine their truth-values. Hence, a case of indeterminacy.

### 4.2.5 Taking Stock

I have argued that Krajewski's Non-Uniqueness Theorem goes against the No-Difference thesis and the claim that arithmetical truth is determinate (suitably understood). In the remainder I propose a new reading of the Theorem and its consequences. In essence, I will take the disagreement between satisfaction classes at face value, as pointing to sentences that are both true and false. Brown (2002, 632) notes that one motivation for believing in sentences that are both true and false is found in the treatment of the logical paradoxes: '*the dialetheists take paradoxes such as the liar and the paradoxes of naïve set theory at face value.* That is, they view these paradoxes as proofs that certain inconsistencies are true' (my emphasis). My approach to mutually inconsistent satisfaction classes shares the same motivation. It is perhaps not that surprising then that the new semantics can be used to engender paraconsistent models of arithmetic. I argue below that these models afford a way to evade the alleged cases of indeterminacy brought about by Krajewski's Theorem.

## 4.3 A New Reading of Krajewski's Theorem

### 4.3.1 The Idea

Since satisfaction classes provide a formal characterization of truth, their extensions should somehow contribute to the truth-value of any given sentence. And if different satisfaction

classes can disagree on a sentence with respect to membership, this fact should again be reflected on that sentence's truth-value. It is then natural to consider a new valuation function defined relative to a model of PA admitting a full satisfaction class, and ranging over the model's different classes. This motivates the following definition:

**Definition 4.3.1** (Multi-Valued Valuation)**.** *For a model $\mathfrak{M}$ of PA admitting a full satisfaction class, and $\mathbb{I}$ an index set for all full satisfaction classes in $\mathfrak{M}$, we define a valuation v relative to the model as a function from (codes) of sentences to truth-values such that $v : Sent_{PA}^{\mathfrak{M}} \rightarrow \{0, 1/2, 1\}$ and:*

$$v(\varphi) = \begin{cases} 1 & \text{if } \varphi \in S_i \text{ for all } i \in \mathbb{I} \\ 0 & \text{if } \varphi \notin S_i \text{ for all } i \in \mathbb{I} \\ 1/2 & \text{otherwise} \end{cases}$$

Since the different $S_i$'s are meant to capture true sentences in $\mathfrak{M}$, the interpretation of the integers should be straightforward. If $\varphi$ is assigned the value 1 then the sentence belongs to the extension of every satisfaction class so that $\varphi$ is interpreted as being 'only true'; similarly, if $\varphi$ is assigned the value 0 then the sentence does not belong to the extension of any satisfaction class so that $\varphi$ is interpreted as being 'only false'. How to interpret 1/2 is a more delicate matter.

### 4.3.2 Interpreting 1/2

If there is disagreement among satisfaction classes, then the sentence can be thought of as true with respect to the interpretation of some predicates and false with respect to others. Again, different satisfaction classes *pull in different directions*. I think that there are plausible interpretations of 1/2 available.

An analogy will help. A predicate like 'heap' is a vague predicate since it admits of borderline cases for which it is indeterminate whether a given collection of grains is a heap or not. Presumably, there is a collection of grains $h$ for which it is not determinate that $h$ forms a heap nor is it determinate that $h$ does not form a heap. But we can think of precisifications of the language

where the borderline cases disappear: on some precisifications the collection $h$ counts as a heap while on others it does not. Modelling logical consequence between sets of formulas containing vague predicates is a complex task. On the supervaluationist (SpV) approach popularised by Fine (1975), semantic consequence is defined as preservation of super-truth —where a formula is super-true if is true on *all* possible precisifications of the vague components occurring in the formula. Here, cases of indeterminacy, as in the statement '$h$ is a heap', are neither (super-)true nor (super-)false. In contrast, the subvaluationist (SbV) approach of for example Hyde (1997, 2010) defines semantic consequence in terms of preservation of sub-truth — where a formula is sub-true if it is true on *some* precisification of the language. In this new case, indeterminacy is no longer understood in terms of truth-value gaps but instead truth-value gluts: the statement '$h$ is a heap' is true under some precisifications and false under others, and so is both (sub-)true and (sub-)false. The theories of both logics are well-known. Some basic facts are summarised in the next proposition (where $\varphi, \psi$ stand for formulas in classical propositional logic CL, and where $\models_{\mathsf{SpV}}$ is a multiple-conclusion consequence relation[9]):

**Proposition 4.3.2** (e.g. Hyde (1997))**.**

1. $\models_{SpV} \varphi$ *if and only if* $\models_{CL} \varphi$ *if and only if* $\models_{SbV} \varphi$.

2. *There are formulas* $\varphi$ *and* $\psi$ *such that* $\varphi \vee \psi \not\models_{SpV} \varphi$ *and* $\varphi \vee \psi \not\models_{SpV} \psi$.
   *There are formulas* $\varphi$ *and* $\psi$ *such that* $\varphi, \psi \not\models_{SbV} \varphi \wedge \psi$.

From (1) we see that SpV and SbV have the same validities and these coincide with classical logic; from (2) that SpV invalidates subjunction while SbV invalidates adjunction. (For an extensive treatment of theories of vagueness the *loci classici* are Keefe (2000) and Williamson (1994).)

I propose to picture the disagreement between satisfaction classes in a similar manner. In a model admitting multiple classes, 'truth' behaves like a vague predicate. Each satisfaction class

[9] If the consequence relation $\vdash$ is a multiple-conclusion relation then $\Sigma \vdash \Gamma$ means that 'if all the members of $\Sigma$ are satisfiable then *some* member of $\Gamma$ is satisfiable'. A multiple-conclusion relation for classical logic, for example, should validate $P \vee Q \vdash P, Q$ since whenever $P \vee Q$ is true it classically follows that either $P$ or $Q$ also is.

is a possible precisification of what it means for a sentence $\varphi$ to be true (in the model) and disagreement entails that $\varphi$ being true is an indeterminate matter. A sentence gets an integer value whenever it does not change in membership across the different satisfaction classes. If the analogy holds, the value 1/2 can be read in two different ways similar to the supervaluationist or the subvaluationist approach, respectively. According to the former, 1/2 is interpreted as a case of 'underdetermination' where disagreement between satisfaction classes renders the sentence 'neither true nor false'. According to the latter, 1/2 is interpreted as a case of 'overdetermination' where disagreement between satisfaction classes renders the sentence 'both true and false'.

But which of these two should we prefer? Underdetermination or overdetermination? In this chapter I propose to study the latter: cases of 1/2 will be cases where a sentence is both true and false.

### 4.3.3 Satisfaction and Consequence

A meta-theoretic notion of semantic consequence aiming to capture truth should take this intermediate value into account *only* on the latter approach. This yields two distinct notions of semantic consequence depending on whether 1/2 is a designated value or not. Keeping the analogy with SpV and SbV, I will call each notion $\models_p$ and $\models_b$ from which we get the following definitions:

**Definition 4.3.3.** *Where $\mathfrak{M}$ is a model of PA admitting a full satisfaction class and $v$ is as presented in Definition 4.3.1, we introduce the following definitions:*

1. *(Supervaluationist Satisfaction $\models_p^{\mathfrak{M}}$) For a model $\mathfrak{M}$ and valuation $v$, we say that '$\varphi$ is satisfied$_p$ in $\mathfrak{M}$' if and only if $v(\varphi) = 1$. If $\varphi$ is satisfied$_p$ we write $\mathfrak{M} \models_p^{\mathfrak{M}} \varphi$.*

   *(Subvaluationist Satisfaction $\models_b^{\mathfrak{M}}$) For a model $\mathfrak{M}$ and valuation $v$, we say that '$\varphi$ is satisfied$_b$ in $\mathfrak{M}$' if and only if $v(\varphi) \geq 1/2$. If $\varphi$ is satisfied$_b$ we write $\mathfrak{M} \models_b^{\mathfrak{M}} \varphi$.*

2. *(Supervaluationist Semantic Consequence$_p$) We say that $\Sigma$ (semantically) entails$_p$ $\varphi$ in $\mathfrak{M}$' if and only if $v(\varphi) = 1$ if $v(\sigma) = 1$ for all $\sigma \in \Sigma$. If $\Sigma$ (semantically) entails$_p$ $\varphi$ we write $\Sigma \models_p^{\mathfrak{M}} \varphi$.*

*(Subvaluationist Semantic Consequence$_b$) We say that '$\Sigma$ (semantically) entails$_b$ $\varphi$ in $\mathfrak{M}$' if and only if $v(\varphi) \geq 1/2$ if $v(\sigma) \geq 1/2$ for all $\sigma \in \Sigma$. If $\Sigma$ (semantically) entails$_b$ $\varphi$ we write $\Sigma \models_b^{\mathfrak{M}} \varphi$.*

Some remarks are in order. On both accounts, semantic consequence still corresponds to preservation of truth. The difference is that with consequence$_p$ a sentence taking the intermediate value is akin to a truth-gap and so is not true. This is in sharp contrast to consequence$_b$ where again such a sentence is a truth-glut. Furthermore, in both cases, semantic consequence is relative to a valuation $v$ which, in turn, is relative to the model and the model's full satisfaction classes — the superscript $\mathfrak{M}$ is there to remind us of this very fact. Unlike the standard notion of logical consequence, in our case, consequence$_{p,b}$ is always relative to a particular model. In essence, this is a by-product of the way we have built $v$: note that $v(\varphi)$ is only defined whenever $\varphi$ is a code of a sentence in the model's sense, and, of course, whether something is a sentence in the model's sense is local to the model. Finally, this last observation points to another way in which our definition deviates from common practice: while standard logical consequence is a binary relation between sets of formulas, ours is a binary relation between sets of *codes* of formulas.

Interestingly, though consequence$_{p,b}$ is dependent on $v$ and so relative to models of PA, both notions mimic the Tarskian clauses for classical semantic consequence in the following sense:

**Proposition 4.3.4.** *Let $\Sigma, \Sigma' \subseteq Sent_{\mathsf{PA}}^{\mathfrak{M}}$. The relation $\models_{p,b}^{\mathfrak{M}}$ has the following properties:*

*Reflexivity: If $\varphi \in \Sigma$ then $\Sigma \models_{p,b}^{\mathfrak{M}} \varphi$*
*Monotonicity: If $\Sigma \models_{p,b}^{\mathfrak{M}} \varphi$ and $\Sigma \subseteq \Sigma'$ then $\Sigma' \models_{p,b}^{\mathfrak{M}} \varphi$*
*Transitivity: If $\Sigma \models_{p,b}^{\mathfrak{M}} \varphi$ and $\Sigma', \varphi \models_{p,b}^{\mathfrak{M}} \psi$ then $\Sigma \cup \Sigma' \models_{p,b}^{\mathfrak{M}} \psi$*

*Proof Idea.* Reflexivity follows from the definition of $\models_p^{\mathfrak{M}}$ and $\models_b^{\mathfrak{M}}$. For Monotonicity under $\models_b^{\mathfrak{M}}$, for instance, suppose $v(\sigma') \geq 1/2$ for all $\sigma' \in \Sigma'$; therefore, $v(\sigma) \geq 1/2$ for all $\sigma \in \Sigma$; from this and from $\Sigma \models_b^{\mathfrak{M}} \varphi$ it follows that $v(\varphi) \geq 1/2$. Transitivity is similar to Monotonicity. □

It can be shown that a statement similar to Proposition 4.3.4 holds whenever a logic is defined in terms of preservation of designated values. (See Button and Walsh (2018, 136–138) and Suszko (1975) for a statement and proof of this more general result.)

### 4.3.4 Subvaluationist Truth

I now want to find some constraints on our notion of semantic consequence $\models_b^{\mathfrak{M}}$ for any model $\mathfrak{M}$ of PA admitting full satisfaction classes. The first thing to note in this regard is that the consequence relation is paraconsistent (or non-explosive) in a very particular sense. To see how, I first need to introduce a distinction between distributive and collective contradictions, and between distributive and collective explosions.

**Definition 4.3.5** (Collective and Distributive Contradiction/Explosion)**.** *Let $\mathcal{L}$ be a first-order language and $\varphi$ an arbitrary well-formed formula in $\mathcal{L}$. A contradiction in $\mathcal{L}$ is distributive if and only if it is of the syntactic form '$\varphi, \neg\varphi$'; it is collective if and only if it is of the syntactic form '$\varphi \wedge \neg\varphi$'. For a logic $\langle \mathcal{L}, \models \rangle$ we call the relation $\models$ collectively explosive if and only if $\varphi \wedge \neg\varphi \models \psi$ and distributively explosive if and only if $\varphi, \neg\varphi \models \psi$, for every $\psi \in \mathcal{L}$.*

We can replicate this distinction in terms of the relation $\models_b^{\mathfrak{M}}$ in the following way:

**Definition 4.3.6** (Collective and Distributive Explosion for $\models_b^{\mathfrak{M}}$)**.** *Let $\varphi, \psi$ be arbitrary sentences in $Sent_{\mathsf{PA}}^{\mathfrak{M}}$. We call the relation $\models_b^{\mathfrak{M}}$ collectively explosive if and only if $\varphi \wedge \neg\varphi \models_b^{\mathfrak{M}} \psi$ and distributively explosive if and only if $\varphi, \neg\varphi \models_b^{\mathfrak{M}} \psi$.*

We now have:

**Proposition 4.3.7.** *$\models_b^{\mathfrak{M}}$ is collectively explosive.*

*Proof.* Let $\mathfrak{M}$ be an arbitrary model of PA admitting a full satisfaction class. Observe that, by the compositional properties of satisfaction classes, for no sentence $\varphi \in Sent_{\mathsf{PA}}^{\mathfrak{M}}$ is there a class $S_i$ such that $\varphi \wedge \neg\varphi \in S_i$. Otherwise $\varphi \in S_i$ and $\varphi \notin S_i$ by clauses (4) and (5) in definition 4.2.1. Hence, for a valuation $v$ based on $\mathfrak{M}$ it follows that $v(\varphi \wedge \neg\varphi) = 0$. It vacuously follows that $\varphi \wedge \neg\varphi \models_b^{\mathfrak{M}} \psi$ for any $\psi$. □

Let us say that a model $\mathfrak{M}$ *admits mutually inconsistent full satisfaction classes* whenever there is some $m$, $n$ and some $\varphi \in \mathrm{Sent}_{\mathsf{PA}}^{\mathfrak{M}}$ such that $\varphi \in S_m$ but $\varphi \notin S_n$.

**Proposition 4.3.8.** *If $\mathfrak{M}$ admits mutually inconsistent full satisfaction classes then $\models_b^{\mathfrak{M}}$ is not distributively explosive.*

*Proof.* By assumption, $\mathfrak{M}$ is such that there is some $m$, $n$ and some $\varphi \in \mathrm{Sent}_{\mathsf{PA}}^{\mathfrak{M}}$ such that $\varphi \in S_m$ but $\varphi \notin S_n$. Therefore, by clause (5) in definition 4.2.1, $\neg\varphi \notin S_m$ but $\neg\varphi \in S_n$ and, by definition of $v$, $v(\varphi) = v(\neg\varphi) = 1/2$. But, by the proof of Proposition 4.3.7, $v(\varphi \wedge \neg\varphi) = 0$. Letting $\psi := \varphi \wedge \neg\varphi$ we get $\varphi, \neg\varphi \not\models_b^{\mathfrak{M}} \psi$. □

Similarly, note that:

**Proposition 4.3.9.** *If $\mathfrak{M}$ admits mutually inconsistent full satisfaction classes then $\models_b^{\mathfrak{M}}$ is not truth-functional.*

*Proof.* Let $v$ be a valuation based on a model $\mathfrak{M}$ admitting mutually inconsistent satisfaction classes. Reasoning in a similar way to the proof of Proposition 4.3.8, there is some $\varphi$ with $v(\varphi) = v(\neg\varphi) = 1/2$. It suffices to observe that $1/2 = v(\varphi \wedge \varphi) \neq v(\varphi \wedge \neg\varphi) = 0$. Furthermore, $1/2 = v(\varphi \vee \varphi) \neq v(\varphi \vee \neg\varphi) = 1$. □

Failure of distributive explosion also shows that $\models_b^{\mathfrak{M}}$ invalidates many other classical inference rules whenever $\mathfrak{M}$ admits mutually inconsistent full satisfaction classes. In particular, letting $\alpha := \varphi$ be as in Proposition 4.3.8 gives a witness for the failure of the adjunction rule: $\alpha, \neg\alpha \not\models_b^{\mathfrak{M}} \alpha \wedge \neg\alpha$. Define further $\beta := \varphi \wedge \neg\varphi$ and $\gamma := \varphi \vee \neg\varphi$, and let the expression $\varepsilon \to \zeta$ abbreviate $\neg\varepsilon \vee \zeta$. Then the following are all invalid in our semantics: $\alpha, \alpha \to \beta \not\models_b^{\mathfrak{M}} \beta$; $\gamma \to \alpha, \neg\alpha \not\models_b^{\mathfrak{M}} \neg\gamma$; $\gamma \to \alpha, \alpha \to \beta \not\models_b^{\mathfrak{M}} \gamma \to \beta$.

I postpone some philosophical comments pertaining to these inferences, or to the failure thereof, to a later section. For now note that all these counterexamples involve instances of sentences taking a middle-value. This makes it natural to think that failures of classical inference rules always involve sentences taking 1/2. For example, the modus ponens rule: is it the case that for any $\varphi, \psi \in \mathrm{Sent}_{\mathsf{PA}}^{\mathfrak{M}}$: $\varphi, \varphi \to \psi \models_b^{\mathfrak{M}} \psi$ whenever $v(\varphi) \neq 1/2 \neq v(\psi)$? This is answered in

the positive. To see why, first augment the language with an operator $\circ$ which is prefixed to sentences in $\text{Sent}_{\text{PA}}^{\mathfrak{M}}$ and that keeps track of when a sentence takes a middle-value (similar to a so-called determinacy operator for vague languages). Its semantics is described by table 4.1. The motivation for the operator is readily seen: $v(\circ\varphi) = 1$ if and only if $v(\varphi) \neq 1/2$ if and only if all satisfaction classes agree on the membership of $\varphi$ and similarly on the membership of $\neg\varphi$.[10] We can now see why our earlier supposition about modus ponens holds.

Table 4.1: $\circ$-Operator

| $v(\varphi)$ | $v(\circ\varphi)$ |
|---|---|
| 1 | 1 |
| 1/2 | 0 |
| 0 | 1 |

**Proposition 4.3.10.** *Let $\mathfrak{M}$ be an arbitrary model of PA admitting a full satisfaction class. We show that*

$$\circ\varphi, \circ\psi, \varphi, \varphi \to \psi \models_b^{\mathfrak{M}} \psi$$

*Proof.* Notice first that the membership of a conditional sentence to a satisfaction class $S_i$ is determined by the membership of its constituents to that predicate, as given by (3) and (5), in a way that mimics the semantics for conditional formulas in classical propositional logic CL. That is, if we say that a sentence $\varphi \in \text{Sent}_{\text{PA}}^{\mathfrak{M}}$ 'scores' 1 whenever it belongs to a satisfaction class $S_i$, and that it scores 0 whenever it does not belong to a satisfaction class $S_i$, then membership of a conditional sentence will be computed by a classical truth-table as follows:

| $\varphi \in S_i$ | $\psi \in S_i$ | $\varphi \to \psi \in S_i$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 1 |
| 0 | 0 | 1 |

[10] Strictly speaking '$v(\circ\varphi)$' is ill-formed since $v$ only takes as arguments sentences in $\mathfrak{M}$. To be precise we should extend $v$ to a new function $v^*$ ranging over formulas in $\text{Sent}_{\text{PA}}^{\mathfrak{M}}$ possibly prefixed with the $\circ$-operator. Since this will not affect the main argument, I will assume that $v$ is itself the required extension.

Suppose that $v(\circ\varphi) = v(\circ\psi) \geq 1/2$. It immediately follows by the semantics of the $\circ$-operator that $v(\circ\varphi) = v(\circ\psi) = 1$ so that the membership of the sentences is decided in the same way across all satisfaction classes. As a result $v(\varphi)$ and $v(\psi)$ take integer values and for an arbitrary satisfaction class $S_i$:

$$\varphi \in S_i \Leftrightarrow \varphi \text{ belongs to every satisfaction class} \Leftrightarrow v(\varphi) = 1$$

A similar claim for $\psi$. Now, let $P$ and $Q$ be propositional variables in classical propositional logic CL. Since $v(\psi)$ and $v(\psi)$ take integer values, we can consider a classical Boolean valuation $f$ from formulas in CL to $\{0, 1\}$ such that $f(P) = v(\varphi)$ and $f(Q) = v(\psi)$. Given the earlier equivalences, we have:

$$\varphi \in S_i \Leftrightarrow f(P) = 1$$

A similar claim for $\psi$ and $Q$. We can now take note of the following equivalences:

$$f(P \rightarrow Q) = 1 \Leftrightarrow f(P) = 0 \text{ or } f(Q) = 1 \Leftrightarrow \varphi \notin S_i \text{ or } \psi \in S_i \Leftrightarrow$$
$$\Leftrightarrow \neg\varphi \in S_i \text{ or } \psi \in S_i \Leftrightarrow \neg\varphi \vee \psi \in S_i \Leftrightarrow \varphi \rightarrow \psi \in S_i$$

Recall that we are assuming that $v(\circ\varphi) = v(\circ\psi) \geq 1/2$ and, therefore, that $v(\circ\varphi) = v(\circ\psi) = 1$. Since the membership of a sentence to a satisfaction class is determined by the membership of its constituents, and since by assumption $\varphi$ and $\psi$ do not change membership across satisfaction classes, it is also the case that $v(\circ\varphi \rightarrow \psi) = 1$. Similar to before, we have:

$$\varphi \rightarrow \psi \in S_i \Leftrightarrow \varphi \rightarrow \psi \text{ belongs to every satisfaction class} \Leftrightarrow v(\varphi \rightarrow \psi) = 1$$

From this and from our earlier equivalences, we finally get:

$$f(P \rightarrow Q) = 1 \Leftrightarrow v(\varphi \rightarrow \psi) = 1$$

Assume now that $v(\varphi) = v(\varphi \rightarrow \psi) \geq 1/2$. Since we know that membership for these formulas is decided in the same way across all satisfaction classes it follows that $v(\varphi) = v(\varphi \rightarrow$

$\psi) = 1$. Suppose moreover that $v(\psi) = 0$. As a consequence, $f$ is a classical valuation satisfying $P$ and $P \to Q$ but not $Q$. But this contradicts the fact that $P, P \to Q \models_{\mathsf{CL}} Q$. Hence, it must be the case that $v(\psi) \neq 0$, or equivalently, $v(\psi) \geq 1/2$.

We conclude that: $\circ\varphi, \circ\psi, \varphi, \varphi \to \psi \models_b^{\mathfrak{M}} \psi$. Failure of the modus ponens rule must involve sentences taking intermediate values.

□

We now generalise this result:

**Proposition 4.3.11.** *Let $\mathfrak{M}$ be an arbitrary model of PA admitting a full satisfaction class. Let $P$ be a collection of propositional variables and $\Sigma(P)$ a set of formulas in $\mathcal{L}_{CL}$ whose variables are all in $P$. Similarly for a formula $\varphi(P)$.*

*Consider a bijection $\pi$ from the collection $P$ to a set $P^* \subseteq Sent_{PA}^{\mathfrak{M}}$. Moreover, for a formula $\psi(P) \in \Sigma(P) \cup \{\varphi(P)\}$, let $\psi(P^*)$ be the sentence just like $\psi(P)$ but where we replace each $P_j$ in $P$ with the code $\pi(P_j) = \alpha_j \in P^*$ and compute the corresponding Boolean combinations. That is, for instance, if $P_1 \wedge P_2 \in \Sigma(P)$ then the corresponding sentence in $\Sigma(P^*)$ will be the number which $\mathfrak{M}$ thinks codes the conjunction of $\pi(P_1)$ with $\pi(P_2)$.*

*It follows that:*

$$\textit{If } \Sigma(P) \models_{CL} \varphi(P) \textit{ then } \{\circ\alpha_j : \alpha_j \in P^*\}, \Sigma(P^*) \models_b^{\mathfrak{M}} \varphi(P^*)$$

*Proof.* Assume $\{\circ\alpha_j : \alpha_j \in P^*\}, \Sigma(P^*) \not\models_b^{\mathfrak{M}} \varphi(P^*)$. Then $v(\circ\alpha_j) \geq 1/2$ and therefore $v(\circ\alpha_j) = 1$ for all such $\alpha_j$. Every satisfaction class will agree with respect to the membership of the $\alpha_j$'s and, therefore, with respect to the membership of each sentence in $\Sigma(P^*)$ and of the sentence $\varphi(P^*)$. It follows that $v(\varphi(P^*)) = 0$ and $v(\sigma(P^*)) = 1$ for all $\sigma(P^*) \in \Sigma(P^*)$. Furthermore, since each $\alpha_j$ takes an integer value, we can consider a classical Boolean valuation $f$ in $\mathsf{CL}$ such that $f(P_j) = v(\alpha_j)$. Because satisfaction classes obey the compositional rules for negation and conjunction as per clauses (4) and (5) of definition 4.2.1 it follows that for any formula $\psi(P) \in \Sigma(P) \cup \{\varphi(P)\}$, $f(\psi(P)) = v(\psi(P^*))$. Therefore, the valuation $f$ is a model for $\Sigma(P)$ but not for $\varphi(P)$. Hence, $\Sigma(P) \not\models_{\mathsf{CL}} \varphi(P)$. This proves the contrapositive. □

However the converse does not hold. We know by 4.2.2 that for any standard sentence $\varphi$: $\mathfrak{M} \models \varphi$ if and only if $v(\ulcorner\varphi\urcorner) = 1$. So take any such arbitrary standard $\varphi \in \mathcal{L}_{\mathsf{PA}}$ (in this case $\varphi$ is being used as a standard sentence, and not a code). Since $\mathfrak{M} \models \varphi \vee \neg\varphi$ it follows that $v(\ulcorner\varphi \vee \neg\varphi\urcorner) = v(\circ\ulcorner\varphi \vee \neg\varphi\urcorner) = 1$ and therefore that $\circ\ulcorner\varphi \vee \neg\varphi\urcorner \models_b^{\mathfrak{M}} \ulcorner\varphi \vee \neg\varphi\urcorner$. But for any propositional letter, say, $P$ replaced by $\ulcorner\varphi \vee \neg\varphi\urcorner$ we have $\not\models_{\mathsf{CL}} P$. This disproves the converse.

Proposition 4.3.11 makes clear the similarity between $\models_b^{\mathfrak{M}}$ and the subvaluationist semantics SbV. For as a special case of this result we obtain the implication:

$$\textit{If } \models_{\mathsf{CL}} \varphi(P) \textit{ then } \{\circ\alpha_j : \alpha_j \in P^*\} \models_b^{\mathfrak{M}} \varphi(P^*)$$

This together with failure of adjunction for models admitting mutually inconsistent full satisfaction classes:

$$\varphi, \psi \not\models_b^{\mathfrak{M}} \varphi \wedge \psi$$

give us a sense in which the properties of $\models_b^{\mathfrak{M}}$, for some models, resemble those of SbV as explained in facts 1 and 2 of Proposition 4.3.2. Proposition 4.3.11 also shows that $\models_b^{\mathfrak{M}}$ in the language extended with the $\circ$-operator validates a weak explosion principle.

**Proposition 4.3.12** (Weak Distributive Explosion)**.** $\circ\varphi, \varphi, \neg\varphi \models_b^{\mathfrak{M}} \psi$

*Proof.* By 4.3.11 and $P, \neg P \models_{\mathsf{CL}} Q$. □

### 4.3.5 Modality and Jaśkowski's Logic

To finish my exposition of the properties of the $\models_b^{\mathfrak{M}}$ relation, I propose a translation between our semantics and the modal propositional logic S5. Passing over some details for the moment, the general idea will be to let $S_i$ define a world $w_i$ built out of the elements of $S_i$ themselves, and where $\varphi$ is true in $w_i$ if and only if $\varphi$ belongs to $S_i$. When $\varphi$ is true according to some satisfaction class then, in the corresponding possible world structure, there is some world where $\varphi$ is true. Since $\varphi$ might be true in any world/satisfaction class we better make sure that the accessibility relation is reflexive (we can always check if in our world $\varphi$ is true) and euclidean (we can always check if in any other world $\varphi$ is true). As a result our possible world structure should define

an S5-structure. This picture points to a connection between truth in S5 and truth in our semantics. More precisely, we have:

**Proposition 4.3.13.** *Let $\mathfrak{M}$ be an arbitrary model of PA admitting a full satisfaction class. Let $P$ be a collection of propositional variables and $\Sigma(P)$ a set of formulas in propositional CL whose variables are in $P$. Similarly for $\varphi(P)$. Let $\pi$ be defined as in 4.3.11. It follows:*

$$\textit{If } \{\Diamond\sigma : \sigma \in \Sigma(P)\} \models_{S5} \Diamond\varphi(P) \textit{ then } \Sigma(P^*) \models_b^{\mathfrak{M}} \varphi(P^*)$$

*Proof Idea.* For each full satisfaction class $S_i$ in $\mathfrak{M}$, let $w_i = \{P_j : \pi(P_j) \in S_i\}$. Now define $\mathfrak{M}_K = \langle W, R, V\rangle$ such that:

$$W = \bigcup_{i\in\mathbb{I}} w_i \qquad R = W \times W \qquad V(P_j) = \{w_i \in W : P_j \in w_i\}$$

Because $R$ defines an equivalence relation, $\mathfrak{M}_K$ is a model for S5. Also note that by construction of $\mathfrak{M}_K$:

$$\mathfrak{M}_K, w_i \models P_j \text{ if and only if } \pi(P_j) \in S_i \quad \text{for any } i,j \tag{*}$$

Suppose that $\{\Diamond\sigma : \sigma \in \Sigma(P)\} \models_{S5} \Diamond\varphi(P)$ holds and assume $v(\sigma(P^*)) \geq 1/2$ for all $\sigma(P^*) \in \Sigma(P^*)$. Then, for each $\sigma(P^*)$, there is one $S_i$ such that $\sigma(P^*) \in S_i$. From (*), for each $\pi(P_j) \in \sigma(P^*)$ it follows that $P_j$ is true in $w_i$ if and only if $\pi(P_j) \in S_i$. Joining this with the fact that satisfaction classes are compositional for conjunction and negation, it follows:

$$\mathfrak{M}_K, w_i \models_{S5} \sigma \text{ if and only if } \sigma(P^*) \in S_i$$

Therefore, for each $\sigma(P^*) \in \Sigma(P^*)$ there is an $i \in \mathbb{I}$ such that $\mathfrak{M}_K, w_i \models_{S5} \sigma$ and, therefore, $\mathfrak{M}_K \models_{S5} \Diamond\sigma$. Using now the fact that $\{\Diamond\sigma : \sigma \in \Sigma(P)\} \models_{S5} \Diamond\varphi(P)$ gives us $\mathfrak{M}_K \models_{S5} \Diamond\varphi(P)$ and, thence, $\mathfrak{M}_K, w_m \models_{S5} \varphi(P)$ for some $w_m$. Again, by (*) and the compositionality of satisfaction classes, we get $\varphi(P^*) \in S_m$. Hence, $v(\varphi(P^*)) \geq 1/2$. This shows that $\Sigma(P^*) \models_b^{\mathfrak{M}} \varphi(P^*)$. □

Proposition 4.3.13 offers a quicker argument for collective explosion as per Proposition 4.3.7; this is because $\Diamond(\varphi \wedge \neg\varphi) \models_{\mathsf{S5}} \Diamond\psi$ where $\varphi$ and $\psi$ are formulas in CL. And note that a similar argument as the one given in relation to the converse of 4.3.11 also proves that in this new case the converse does not hold either.

Curiously, SbV also has a modal interpretation, inspired by Jaśkowski's ([1948] 1969) discussive (also called 'discursive') logic $\mathsf{D}_2$. The motivation for $\mathsf{D}_2$ was to model the total sum of information generated from a set of different sources possibly in disagreement. Call this sum a 'discussive system' ([1948] 1969, 149). Jaśkowski pictured each source in a discussive system as a possible world in an S5-model with the sentences true-at-a-world representing the information that that possible world/source contributed to the overall sum. If $\Diamond\alpha$ is globally true in the model then $\alpha$ is true at some possible world/source and, therefore, true in the discussive system. Given this, Jaśkowski goes on to define 'discussive consequence' via an interpretation in S5 in the following way: $\Sigma \models_{D_2} \alpha$ if and only if $\{\Diamond\sigma : \sigma \in \Sigma\} \models_{\mathsf{S5}} \Diamond\alpha$. (For a survey see also Berto, 2007, chapter 6.) The logic is interesting for invalidating classical inference rules like modus ponens, adjunction and distributive explosion. To see why it suffices to note that in S5 the formulas in each rule prefixed with a possibility operator do not yield validities: for example, $\Diamond\alpha, \Diamond(\alpha \to \beta) \not\models_{\mathsf{S5}} \Diamond\beta$.

Let each precisification induce a possible world in an S5 model where the accessibility relation is total. We can define a formula to be sub-true in SbV if and only if it is true in some world in the resultant possible world model if and only if it is possibly true in every world. In this sense, the logic SbV describes a discussive system like Jaśkowski's (for elaboration see Hyde, 1997), so that proposition 4.3.13 gives us another way in which $\models_b^{\mathfrak{M}}$ resembles the properties of SbV.

### 4.3.6 Taking Stock

We have introduced a new relation of logical consequence $\models_b^{\mathfrak{M}}$ defined relative to each model $\mathfrak{M}$ of PA admitting a full satisfaction class. We have found that (1) $\models_b^{\mathfrak{M}}$ is collectively but not distributively explosive for some models, (2) the valuation is not truth-functional for some mod-

els, (3) any counterexamples to classical validities involve sentences whose membership changes across different satisfaction classes, and, finally, (4) there is a connection between S5-consequence and consequence$_b$. As we expected, connecting the phenomenon of mutually inconsistent satisfaction classes with the existence of sentences that are both true and false generates a system similar to the subvaluationist logics for vague languages. In the next section I provide some philosophical commentary on these constructions.

## 4.4 Philosophical Commentary

### 4.4.1 $\mathsf{D}_2$, Non-Adjunction and Modality

As the discussion about S5 has already made clear, in a way, the subvaluationist-like consequence $\models_b^{\mathfrak{M}}$ resembles Jaśkowski's ([1948] 1969) discussive logic $\mathsf{D}_2$. Discussive logic however has been charged with a non-standard account of conjunction as well as a 'modalised' approach to contradictions, allegedly rendering the system unnatural. Given the affinity between $\mathsf{D}_2$ and $\models_b^{\mathfrak{M}}$, it is worth checking whether the complaints about $\mathsf{D}_2$ affect our case.

First, note that with $\models_b^{\mathfrak{M}}$ distributive explosion is avoided only at the cost of dropping the adjunction rule. As per 4.3.8, for some models, there is some non-standard $\varphi$ for which $\varphi, \neg\varphi \not\models_b^{\mathfrak{M}} \varphi \wedge \neg\varphi$. But against non-adjunctive approaches it is common to object that the rule is analytic of the meaning of conjunction: that the meaning of the connective is exhausted by an equivalence of the form $\alpha, \beta \dashv\vdash \alpha \wedge \beta$. Accordingly, a connective $\Delta$ such that $\alpha, \beta \nvdash \alpha\Delta\beta$ is not *real* conjunction.

However, we should be careful not to overstate the problem. After all, it is to be expected that changes in the truth-value of some formulas will come to be reflected in the logical inferences between those formulas themselves. This is no more surprising than when a paraconsistent semantics invalidate disjunctive syllogism. More importantly, in the case of $\models_b^{\mathfrak{M}}$, failure of adjunction is restricted to inferences between sets of non-standard sentences, or codes thereof. Equivalently, any inferences between sets of standard sentences agrees with the rule. This can be seen as follows.

**Proposition 4.4.1.** *Let $\varphi$ be a standard sentence in $\mathcal{L}_{PA}$. It follows that $PA \models \varphi$ if and only if $v(\ulcorner\varphi\urcorner) = 1$ for all models $\mathfrak{M} \models PA$ admitting a full satisfaction class and respective valuation $v$. Similarly, $PA \models \neg\varphi$ if and only if $v(\ulcorner\varphi\urcorner) = 0$ for all models $\mathfrak{M} \models PA$ admitting a full satisfaction class and respective valuation $v$.*

*Proof.* I prove the first claim, the other case being identical.

($\Rightarrow$) Assume $\mathsf{PA} \models \varphi$. Let $\mathfrak{M}$ be a model of PA admitting a full satisfaction class $S_i$. It follows that $\mathfrak{M} \models \varphi$. Since $\varphi$ is standard, by 4.2.2, $\ulcorner\varphi\urcorner \in \bigcap_{i\in\mathbb{I}} S_i$ and so $v(\ulcorner\varphi\urcorner) = 1$. Since $\mathfrak{M}$ was arbitrary the claim holds.

($\Leftarrow$) Assume $\mathsf{PA} \not\models \varphi$. Then there is some model $\mathfrak{M}$ of PA such that $\mathfrak{M} \models \neg\varphi$. By the Löwenheim-Skolem theorems we can take this model to be countable. By theorem 4.2.7, $\mathfrak{M}$ has a countable computably saturated elementary extension $\mathfrak{A}$. Therefore, $\mathfrak{A} \models \neg\varphi$. Moreover, by theorem 4.2.4, $\mathfrak{A}$ admits a full satisfaction class $S_i$. Since $\varphi$ is standard, by 4.2.2, $\ulcorner\varphi\urcorner \notin \bigcup_{i\in\mathbb{I}} S_i$ and so $v(\ulcorner\varphi\urcorner) = 0$. This proves the claim. □

Now, if $\Sigma$ is a collection of standard sentences, let $\ulcorner\Sigma\urcorner := \{\ulcorner\sigma\urcorner : \sigma \in \Sigma\}$. A similar argument to 4.4.1 proves that $\models_b^{\mathfrak{M}}$ agrees with the classical inference rules for the standard part of the language. That is:

**Proposition 4.4.2.** *Let $\Sigma \cup \{\varphi\}$ be a collection of sentences in $\mathcal{L}_{PA}$.*
*It follows that $PA \cup \Sigma \models \varphi$ if and only if $\ulcorner\Sigma\urcorner \models_b^{\mathfrak{M}} \ulcorner\varphi\urcorner$ for all models $\mathfrak{M} \models PA$ admitting a full satisfaction class and respective valuation $v$.*

Propositions 4.4.1 and 4.4.2 illustrate the important idea of 'classical recapture'. It is thought that any theory of inconsistent arithmetic has to allow for a recapture of the classical validities in the sense that some suitable part of the classically valid formulas should remain true or satisfiable in the models of the inconsistent theory. Ignoring the ambiguity over what is to count as a 'suitable part' of the classical validities, the relevant point is that, by preserving classical inference rules involving the standard part of the language, $\models_b^{\mathfrak{M}}$ already allows for a substantial recapture of classical arithmetic. This means that $\models_b^{\mathfrak{M}}$ does not presuppose a radical revision

of classical inferences nor is, thereby, rendered unnatural. Where it matters, the standard part, adjunction is preserved.

Besides, the relation between analyticity and adjunction is disputable at best. Sure, we may grant for the sake of argument that, in ordinary contexts, usages of a non-adjunctive 'and' are infelicitous. Nonetheless, it would be hasty to conclude that the rule is analytic for conjunction. This is because there is genuine disagreement when it comes to the rule. Not only Jaśkowski ([1948] 1969) and Hyde (1997), but also Belnap (1977) imagines a scenario where a non-adjunctive conjunction is preferred. Very generally, Belnap imagines a computer program that returns 'Yes' to a query if and only if the relevant content has been introduced in the computer's data bank. If I enter $\varphi$ then the computer will return 'Yes' to a query about $\varphi$ and similarly if the reader enters $\psi$ the computer will return 'Yes' to a query about $\psi$. But the computer will only return 'Yes' to $\varphi \wedge \psi$ if and only if both $\varphi$ and $\psi$ come from the same source. Therefore, if $\psi$ is just $\neg\varphi$ then the computer may return 'Yes' to both $\varphi$ and $\neg\varphi$ and return 'No' to their conjunction. Indeed, the model is reminiscent of Jaśkowski's discussive system. Genuine disagreement over the rule, among competent speakers and logicians like the ones just cited, should give us pause when it comes to rash verdicts about what is to count as analytic.

So much for analyticity. For a different kind of worry, we might complain that the way that $\models_{\flat}^{\mathfrak{M}}$ makes some sentences both true and false is too 'modalised'. To better understand the problem let us first state the original objection to $\mathsf{D}_2$. As we saw, a formula is true in a discussive system if it is true in some possible world in an S5-structure. However, when different sources are in disagreement with respect to $\alpha$, it might be the case that $\alpha$ holds in one world while $\neg\alpha$ holds in another so that $\alpha$ and $\neg\alpha$ are both true in the system. Failure of distributive explosion is achieved through a kind of modalisation of formulas. A distributive contradiction is true not by there being a unique formula both true and false — a formula both true and false in *one* world — but by the overall contribution of different formula's truth-values in *different* worlds. On the contrary, were $\alpha \wedge \neg\alpha$ true in the discussive system then any $\beta$ would be true as well in virtue of $\Diamond(\alpha \wedge \neg\alpha) \models_{\mathsf{S5}} \Diamond\beta$. And for Priest and Routley this is a disappointing feature of the logic:

> The main problem with the discursive approach is just that it does not take the second, dialethic, motivation (that there are true contradictions) seriously. Contradictions "may be true" but this amounts to no more than "true in different worlds". Moreover each possible world is as consistent as any classicist could wish: the approach is too much modally based to accommodate inconsistency satisfactorily. (Priest & Routley, 1989, 162)

True contradictions in $\models_b^{\mathfrak{M}}$ are also produced through a similar modalisation: a non-standard sentence and its negation belong to *different* satisfaction class. Besides, satisfaction classes themselves are consistent (by definition, $\varphi \in S$ if and only if $\neg\varphi \notin S$) and obey the classical rules for compositional truth. This suggests that a similar objection against $\models_b^{\mathfrak{M}}$ is forthcoming. But what exactly is the objection? Priest and Routley's original remarks are rather perplexing. After all, one would expect that $\alpha$ and $\neg\alpha$ being both true was all that was required for taking the idea that there are true contradictions *seriously*. Apparently, it is not. Is the extra requirement that $\alpha \wedge \neg\alpha$ should be true as well? Or, in a related manner, that the logic should block both distributive and collective explosion? If so, the objection amounts to little more than the observation that $\mathsf{D}_2$ or $\models_b^{\mathfrak{M}}$ collectively explode. Without further elaboration, it is difficult to see what Priest and Routley's objection actually amounts to.

### 4.4.2 Paraconsistent Arithmetic

An interesting feature of $\models_b^{\mathfrak{M}}$ is that it affords a new perspective on how inconsistencies arise within arithmetic. We saw in Chapter 2 how to build inconsistent models of PA in the Logic of Paradox LP. There, from a classical model of PA and from an equivalence relation defined over the model's domain, we considered quotient algebras where the properties of the equivalence classes in the quotient were just the collection of the properties their members had in the original structure — even when those properties were mutually inconsistent. If the original structure was an LP-model of PA then so was the quotient.

The technique is particularly interesting in that it produces finite and decidable models of PA, evading standard limitative results like Church or Gödel's Theorems (see Priest, 1994a, 346–

347 for elaboration). The inconsistencies can start at any point in the natural line. That is, for any $n \in \mathbb{N}$ there is a model of PA where $n$ is identical with any larger number but not identical with any smaller number. Graphically, the successor function in these structures will depict a copy of an initial segment of the naturals with the function looping at the end. This immediately invites a curious question: what is a natural starting point for the loop? What is the first $n$ equal (and also different) to its successor? Well, it is argued that the inconsistencies can be located in such a way that all our everyday arithmetical operations will involve numbers that belong only to the consistent initial segment of the natural line. In particular, Priest goes on to locate the least inconsistent number in an unimaginably faraway point; he writes: '[l]et us, henceforth, fix $n$ as some incredibly large number, say a number larger than the number of combinations of fundamental particles in the cosmos, larger than any number that could be sensibly specified in a lifetime, so large that it has no physical meaning or psychological reality' (1994a, 338).

The models generated by $\models_b^{\mathfrak{M}}$ offer a new perspective on the problem of where to locate the inconsistencies. Under $\models_b^{\mathfrak{M}}$, any both true and false sentences will be encoded by elements belonging to the non-standard part of the model. The place where to locate the inconsistencies in this case is not after some number $n$ but, in a way, in the non-standard part of the language.

### 4.4.3 Revisiting the No-Difference Thesis

Finally, I want to quickly revisit the themes with which we started. In particular, I want to ask how the new reading of Krajewski's Theorem bears on the No-Difference Thesis. To make progress on this issue, observe that determinacy is usually equated with the fixed possession of a *classical* truth-value, or lack thereof. For example, Maddy offers us a quick argument for set-theoretic determinacy: from realism about a unique universe of sets, Maddy directly infers that "CH or not-CH' is a theorem [...] as a fact about $V$; therefore CH is either true or false there' (2011, 63–64). What is noteworthy about argument's of this kind is how few alternatives they leave open. From realism about a unique universe, we are to infer not only that each set-theoretic sentence has a determinate truth-value but, moreover, that each is either determinately

true or determinately false. And we are to draw a similar moral in the case of the plenitudinous platonist: if an independent sentence is decided in different ways across different clusters, it becomes automatically indeterminate. Under plenitudinous platonism, there is no extra truth-value $t$ such that an independent sentence might be determinately $t$.

Nonetheless, intuitively, the problem of arithmetical determinacy is about whether the relevant facts of the matter fix a truth-value for a given sentence. Understood in this way, the concept is silent on the range of truth-values that a sentence might take. And once we allow for both true and false sentences it becomes natural to equate determinacy with the possession of a fixed *classical or non-classical* truth-value. If the preceding is correct, the phenomenon of mutually inconsistent satisfaction classes admits of an altogether different reading from the one we saw.

To repeat. Satisfaction classes can be regarded as collections of true sentences in a model's sense, and so they should play a role in our judgments about truth and determinacy. I suggested earlier that disagreement among satisfaction classes was to be read as evidence of indeterminacy. This was formulated in terms of:

> **Non-Standard Indeterminacy Definition:** For $\mathfrak{M}$ an intended model and $\varphi \in Sent_{\mathsf{PA}}^{\mathfrak{M}}$ a (possibly non-standard) sentence, if $\mathfrak{M}$ admits full satisfaction classes $S_1$ and $S_2$ such that $\varphi \in S_1$ and $\varphi \notin S_2$ then $\varphi$ is indeterminate for $\mathfrak{M}$.

Once we recognise that disagreement among satisfaction classes might be instead evidence of a truth-glut, the Non-Standard Indeterminacy Definition seems to lose some of its intuitive appeal. In essence, the Definition states that there are sentences without a fixed truth-value whenever satisfaction classes disagree. The subvaluationist reading of Krajewski's Theorem suggests, in contrast, that there are sentences with a fixed non-classical truth-value whenever satisfaction classes disagree. Under the subvaluationist approach, the Non-Standard Indeterminacy Definition should be replaced with:

> **Non-Standard Determinacy Definition:** For $\mathfrak{M}$ an intended model and $\varphi \in Sent_{\mathsf{PA}}^{\mathfrak{M}}$ a (possibly non-standard) sentence, if $\mathfrak{M}$ admits full satisfaction classes $S_1$

and $S_2$ such that $\varphi \in S_1$ and $\varphi \notin S_2$ then $\varphi$ is determinately true-and-false for $\mathfrak{M}$.

The new definition yields a sufficient condition for a sentence being determinate, taking into account the idea that both non-standard sentences and satisfaction classes should be taken seriously when evaluating our claims to determinacy. Moreover, it makes clear that determinacy cannot be explained simply as agreement among intended models, that is: a sentence is determinate *if and only if* it is either satisfied by all intended models or falsified by all intended models. The left-to-right condition will have to be dropped. Finally, in this case, the seeming counterexample to the No-Difference thesis is dissolved almost by *fiat*. According to No-Difference, if there is indeterminacy in the truth-value of an arithmetical sentence then there is indeterminacy in the underlying model. As stated, the thesis is compatible with an account of determinacy as possession of possibly non-classical truth-values. Now, let $\mathfrak{M}$ be an intended model of PA and $S_1$ and $S_2$ be full mutually inconsistent satisfaction classes. Let $\psi \in \mathrm{Sent}_{\mathsf{PA}}^{\mathfrak{M}}$ be a sentence such that $\psi \in S_1$ but $\psi \notin S_2$. Since both $S_1$ and $S_2$ are adequate satisfaction classes, it seems that we can equally think of $\psi$ as a true sentence in $\mathfrak{M}$, using $S_1$, and as a false sentence, using $S_2$. But, of course, this does not lead to indeterminacy over the truth of $\psi$. It shows at most that $\psi$ is determinately true-and-false as per the Non-Standard Determinacy Definition. No counterexample is in sight.

## 4.5 Conclusion

The subvaluationist approach restores the No-Difference thesis, and shows how we can make sense of sentences having non-classical truth-values and whose truth is grounded in the disagreement among satisfaction classes. For all the reasons sketched, I think that this new reading of Krajewski's Non-Uniqueness Theorem, the semantics $\models_{b}^{\mathfrak{M}}$ as well as the new way of understanding the concept of 'determinacy' in terms of the possession of a fixed classical or non-classical truth-value are all worthy of further discussion.

# Chapter 5

# Is Strict Finitism Arbitrary?

**Abstract**

Strict finitism posits the existence of a largest natural number. One objection to strict finitism comes from the alleged unacceptable arbitrariness surrounding talk about the largest number. After all, there is no apparent reason as to why the numbers should stop at a specific point and not a bit later in the natural line. Drawing on how arbitrariness arguments are employed in mereology, I propose several ways of understanding this arbitrariness charge against strict finitism. No matter how the charge is read, I argue that it is always found wanting.

## 5.1 Introduction

Strict finitism, as I will be understanding it here, rejects both an actual and potential infinite in arithmetic, positing instead the existence of at most finitely many natural numbers together with a largest natural number. It is a realist view about existential quantification in arithmetic even though it rejects, at the same time, the existence of domains of infinitely many numbers. Strict finitists agree that small initial segments of the classical infinite models of arithmetic correctly describe a portion of mathematical reality. They diverge from orthodoxy, however, by postulating the existence of a largest number, reached after finitely many iterations of the successor function. Depending on the motivation, strict finitists may go on to say that the largest

number's successor is undefined, or even that it is equal to itself (so that the successor function sort of 'loops' at the end).

One objection to strict finitism comes from the alleged unacceptable arbitrariness surrounding talk about the largest number. After all, there is no apparent reason as to why the numbers should stop at a specific point and not a bit later in the natural line, nor as to why the largest number should bear some arithmetical properties instead of others like, say, being even and not odd. I think that these complaints are in the mind of many authors when pondering on strict finitism's shortcomings, and it is the difficulty in providing adequate responses that makes them look so decisive against the view. Call this *the arbitrariness charge*. Even though it is a natural objection to advance, it is also, to my knowledge, never explicitly discussed in the literature (the only exception seems to be Van Bendegem (2012) which I reference below). Be that as it may, I think that the charge's appeal is only apparent. In this chapter, drawing on how arbitrariness arguments are employed in discussions about mereology, I propose three ways of reading this arbitrariness charge against strict finitism: in terms of strict finitism's ontological commitments (section 5.2), in terms of facts of the matter about whether a given number is the largest (section 5.3), and in terms of epistemic principles governing theory choice (section 5.4). No matter how the charge is understood, I argue that it is always found wanting.

## 5.2 Arbitrariness and Ontological Commitment

Arbitrariness arguments are a staple in the literature on mereology. As it is well-known, mereological universalism posits that every collection of objects composes a further object corresponding to their fusion. Under this account there are not only trees and fishes but also extra objects composed of trees and fishes — the 'trishes'. One hugely popular argument for mereological universalism comes from the alleged unacceptable arbitrariness surrounding more restrictive views. (See *inter alia* Hawthorne (2006, vii), Sosa (1999, 178), Van Cleve (1986, 145) and Yablo (1987, 307).) After all, there is no apparent reason as to why there should be trees and fishes but not trishes. Therefore, the argument goes, on pain of arbitrariness, an ontology of trees and

fishes has in addition to include trishes as well as many other fusions. As Korman explains (2020, section 2.5), 'to think otherwise would be to take the facts about what exists to be arbitrary in a way that they plausibly are not'.

Operative in this argument seems to be a principle of domain expansion of the form: for two entities $e$ and $e'$ bearing no ontologically significant difference between each other, if a theory $\mathsf{T}$ is not arbitrary, then if $\mathsf{T}$ is ontologically committed to $e$, then $\mathsf{T}$ is further ontologically committed to $e'$. Contrapositively, for two entities $e$ and $e'$ bearing no ontologically significant difference between each other, if a theory $\mathsf{T}$ is ontologically committed to $e$ but not to $e'$, then $\mathsf{T}$ is arbitrary. (See also Korman (2010, 123–124) for a similar idea.) I think it is this principle, or something fairly similar, which grounds the claim that on pain of arbitrariness an ontology of trees and fishes has in addition to include trishes. Of course, the principle is subject to refinement given that there is no agreed criterion of ontological commitment (Bricker, 2016). On this matter let us just say that a theory comes with truth-conditions, and that a theory is ontologically committed to some entity whenever that entity must exist for the theory to be true. For my purposes, this rough sketch will suffice.

If this is how arbitrariness arguments are understood in mereology at large then one promising strategy interprets the arbitrariness charge against strict finitism in the same manner. According to strict finitism, there is a largest natural number. This is readily formalised in the language of first-order Peano Arithmetic $\mathcal{L}_{\mathsf{PA}} = \{0, S, \oplus, \otimes\}$ in terms of the sentence $\exists x \forall y \exists z (x = y \oplus z)$. Let $c$ be a witness for the sentence. That is, it is a (Skolem) constant whose value is the largest natural number in the structure which the strict finitist takes to be a plausible or correct model of arithmetical reality. (Below I elaborate on what exactly these structures look like.)[1] The number $c$ must exist for the theory of strict finitism to be true. Therefore, strict finitism is ontologically committed to $c$. But there is no ontologically significant difference be-

[1] Indeed, in the absence of a transparent semantics it is not obvious whether the claim 'there is a largest number' can be so easily expressed in terms of $\exists x \forall y \exists z (x = y \oplus z)$. In Peano Arithmetic, for instance, $\oplus$ is usually a total function. But it seems that whenever $\overline{n}$ refers to the largest number then $\overline{n} \oplus \overline{1}$ should no longer be defined. Again, how the sentence $\exists x \forall y \exists z (x = y \oplus z)$ expresses that there is a largest number will be made clearer below in relation to our discussion on finite models of arithmetic. In the present section, however, an 'intuitive' understanding of these expressions will suffice.

tween $c$ and its successor $S(c)$. Now, an instance of the principle of domain expansion gives us: if strict finitism is not arbitrary, then if strict finitism is ontologically committed to $c$, then strict finitism is further ontologically committed to $S(c)$. Nonetheless, under strict finitism, $c$ is the largest number so that there is no $S(c)$ to begin with. It follows that strict finitism is not ontologically committed to $S(c)$. The last clause of the principle of domain expansion — that is, the nested conditional's consequence — is false and, as a result, strict finitism is arbitrary.

This a neat little argument, clarifying the reasons behind strict finitism's alleged arbitrariness: strict finitism is arbitrary because it is not ontologically committed to entities like $S(c)$. Suppose for sake of argument that this reading of the arbitrariness charge is successful. Now, even conceding this point, it is not clear why strict finitism should be *unacceptably* arbitrary? Sider (2007, 224) reminds us that 'there's more to epistemic life than minimizing arbitrariness'. Perhaps then, strict finitism's remaining epistemic credentials will still make it preferable over other views about arithmetic. On the contrary, in this case, strict finitism's arbitrariness would do nothing to move the needle since those other views would be equally arbitrary. To see why, call maximalism the view that everything that can exist, does exist. Arguably, if the goal of avoiding arbitrariness pushes us to more and more expansive ontologies, including trishes and $S(c)$, then it also leads to maximalism. As Sider notes:

> Maximalism is tempting (to the degree that it is) because it minimizes arbitrariness. If maximalism is false, and some consistent objects are present while others are missing, there's a why-question without an answer: why do these objects, but not those, exist? Whereas if maximalism is true, we have a nicely rounded picture of the world, and fewer why-questions go unanswered. Maximalism is attractive for the same reason that plenitudinous views about material ontology are attractive. (Sider, 2007, 224)

Put in terms of the principle of domain expansion, it then seems that there is no ontologically significant difference which justifies ontological commitment to one entity at the same time that precludes commitment to some other. Therefore, arbitrariness constraints dictate the inclusion of any possible entity in our ontology. But note that the relevant entities do not only

comprise material objects and their fusions, but include mathematical objects as well. (Compare with Fairchild (2022, 622): 'Maximalism thus glossed purports to guarantee not just an abundance of bizarre fusions and coincidents, but also (depending on the details) an abundance of mathematical objects [...]'.) Hence, we are forced to include many mathematical objects which popular accounts of arithmetic do not normally accept.

To give an example, take some run-of-the-mill mathematical realism according to which there are countably infinitely many natural numbers exhibiting the structure of the standard model of Peano Arithmetic (PA). Since there does not seem to be any ontologically significant difference between these numbers and the objects of an uncountable model of PA, arbitrariness constraints dictate the inclusion of those other uncountably many objects in our ontology. Hence, standard mathematical realism is arbitrary. Indeed, at the limit, the only non-arbitrary view which is also ontologically committed to the existence of natural numbers seems to be plenitudinous platonism (as in Balaguer, 1998) since it is realist towards the elements of *any* model of arithmetic. All this to say that strict finitists should not feel particularly embarrassed if their view turns out to be arbitrary. Though not definite, the preceding makes clear that much more needs to be said if this reading of the arbitrariness charge is meant to pose a problem to and only to strict finitism. But is the arbitrariness charge even sound?

Well, first, there is a ready reply to the objection from the last paragraph on the part of the run-of-the-mill mathematical realist. This is because it is normally thought that mathematical objects are necessary objects: if some mathematical object exists, then it exists necessarily; if it does not, again, this is so out of necessity. Now, suppose we are realists towards the countably infinitely many natural numbers of the standard model, but eschew the extra uncountably many numbers of the uncountable model. It seems that there is an ontologically significant difference between the numbers that the realist takes to exist and the ones she does not: the realist's numbers exist necessarily, while the others do not. In this case, since the relevant entities bear an ontologically significant difference, there is no violation of the principle of domain expansion. (Indeed, it is relevant to note that Sider's argument for maximalism is, by construction, *only* an argument for the inclusion of objects that *can* exist). Hence, *if run-of-the-mill mathemat-*

*ical realism is true*, then it is not arbitrary. But of course this argument cuts both ways. The necessity of mathematical objects is not particular to mathematical realism and can be adopted by the strict finitist. By parity of reasoning, it follows that *if strict finitism is true*, then it is not arbitrary.

Furthermore, there is a sense in which the arbitrariness charge begs the question. For, as I noted at the start, strict finitism might assume that the largest number's successor is undefined, so that the term '$S(c)$' does no longer refer. If so, it is no longer clear how we should understand certain claims like 'if strict finitism is not arbitrary, then if strict finitism is ontologically committed to $c$, then strict finitism is further ontologically committed to $S(c)$' instrumental in the arbitrariness charge. Better put, assuming that the assumptions in the arbitrariness charge are intelligible comes pretty close to assuming that the expression '$S(c)$' refers and, therefore, it comes pretty close to already rejecting strict finitism. But recall that there was another option. I also noted that strict finitism might instead introduce a loopy successor function, so that $c = S(c) = SS(c) = \ldots$ In this case, the strict finitist has a perfectly good reply to the arbitrariness charge, namely: the premise 'strict finitism is not ontologically committed to $S(c)$' is false, meaning that the whole argument is unsound. This is because if strict finitism is ontologically committed to $c$ then it is equally committed to $S(c)$ in virtue of identifying $c = S(c)$. The strict finitist's correct model of arithmetic reality (in a language augmented with $c$) will satisfy both $\forall y \exists z(c = y \oplus z)$ and $\forall y \exists z(S(c) = y \oplus z)$. For these reasons, I think that this reading of the arbitrariness charge can be resisted.

## 5.3 Arbitrariness and Facts of the Matter

Interestingly, a closer look at mereological universalism yields a different version of the arbitrariness charge. This is because, in mereology, arbitrariness arguments are only one horn of a dilemma against restricted views on composition. If composition is restricted, if some collections have mereological fusions while others do not, then two options usually follow: either there is a sharp cut-off between cases where composition occurs and cases where it doesn't, or

there are indeterminate cases of composition. These two options can be clarified in terms of talk of 'facts of the matter'. In the former case there is always a fact of the matter about whether composition occurs and, therefore, about whether composite entities exist — here the problem is that facts about what exists change so drastically with respect to entities bearing no ontologically significant difference between each other, that they start looking unacceptably arbitrary. In the latter case, sometimes, there are no facts of the matter about whether composition occurs nor whether some composite entities exist. The classic rebuttal of indeterminate composition comes from Lewis (1986, 211–213) who basically finds such discourse unintelligible. As Bennett (2015, 253) summarizes, for Lewis 'there are no vague objects, which entails that existence is not vague, which entails that composition is not vague, which entails that composition is not restricted in any way' (for a defence of indeterminate objects see also Barnes & Williams, 2011).

Still, what is important for strict finitism is that we are clear on the difference between the thesis that there are always facts of the matter about composition, on the one hand, and the thesis that composition is sometimes indeterminate, on the other. To better appreciate the distinction just consider the sorites paradox and imagine a closed series of increasingly bigger collections of grains of sand just that the very first collection in the series is not a heap while the last is a heap. There is a difference between saying that there is a sharp boundary between heaps and non-heaps — such that for each collection of grains in the series there is a fact of the matter about whether that collection is a heap or not —, and that for some collections in the series there is no fact of the matter about whether they are heaps.

Continuing the analogy with the cases from mereology and the sorites, the arbitrariness charge can perhaps be read as the claim that, under strict finitism, for some number, it is indeterminate whether that number is the largest— i.e. there is no fact of the matter about whether that number is the largest. This new reading is also in keeping with our initial characterization of the arbitrariness charge, recall: 'that there is no apparent reason as to why the numbers should stop at a specific point and not a bit later in the natural line'. In the present case, absence of apparent reasons is interpreted in terms of facts of the matter so that, for some $n$, that there is no apparent reason as to why $n$ is the largest number just means that there is no relevant fact

of the matter about whether $n$ is the largest number.

Notice that this is a *de re* claim and not a *de dicto* one. More precisely, it is indeterminacy applied to an object and not to a sentence. Still more precisely, it is the claim that there is some $n$ such that there is no fact of the matter about whether $n$ is the largest number, as opposed to the claim that there is no fact of the matter about whether there is some $n$ such that $n$ is the largest number. Indeed, it is worth noting that this distinction is not strange to the literature on the sorites: let $\sigma(x)$ abbreviate 'a collection of $x$-many grains of sand is not a heap and a collection of $x + 1$-many grains of sand is a heap';[2] according to the supervaluationist solution to the sorites paradox, it is determinate (super-true) that $\exists x\sigma(x)$ even though there is no particular $e$ in the range of the quantifier for which $\sigma(\bar{e})$ is determinate, where $\bar{e}$ is the name for $e$ (see for example Hyde & Raffman, 2018, section 3.3 for elaboration).

Nonetheless, if the arbitrariness charge is understood in this way then it admits of a ready answer. Simply put, the strict finitist's mathematical ontology provides the relevant facts of the matter about whether every number is the largest number or not. To see how, suppose first that the strict finitist identifies the natural numbers with *abstracta*, interpreting arithmetical sentences as relations among *bona fide* objects. Let $n$ be the some number that the strict finitist takes to exist, and $\overline{n}$ its numeral. In this case, the statement '$\overline{n}$ is the largest number' is read as asserting the existence of an abstract object — the reference of '$\overline{n}$' — which bears the relation 'being larger than' to all the other numbers. Hence, there are facts of the matter about whether $n$ is the largest number, namely, the structure of the abstract realm and the existence or lack thereof of an abstract object bearing the relation just sketched. Now, suppose instead that the strict finitist assumes that there are at most finitely many *concreta* and identifies (a sub-collection of) them with the natural numbers, reinterpreting arithmetical statements as relations among concrete objects referred to by numeral-terms. In this case too, there are facts of the matter about whether $n$ is the largest number, namely, the structure of the concrete realm and the relations that hold among the relevant concrete objects. Hence, I also think that this reading of

[2] For any elements $n, m \in \mathbb{N}$, '$n + m$' refers to the classical value of the addition of $n$ and $m$ in the structure $\mathbb{N} = \langle \mathbb{N}, 0, S, +, \times \rangle$.

the arbitrariness charge can be resisted.

## 5.4 Arbitrariness and Theory Choice

### 5.4.1 The Charge

Fairchild (2022) provides a different reading of arbitrariness arguments in connection to mereological universalism. This time the stress is laid on epistemology as opposed to metaphysics or facts about what exists. We are apt to call a choice or a theory arbitrary whenever it involves preferring one between two options epistemically on a par. Moreover, it is plausible to demand suspension of belief when a rational agent is confronted with two equally good theories; an agent who adopts one theory in detriment of another with the same epistemic credentials is, at least at first sight, guilty of a failure of rationality. Given this, Fairchild (2022, 625) goes on to introduce an Epistemic Parity Thesis (adapted in turn from McSweeney, 2019) which is defined as:

> **Epistemic Parity.** For theories T and T', if we have no reasons to favour T over T', and T and T' are incompatible, we should not believe T and we should not believe T'.

As I understand it, the idea now is that a theory T is arbitrary whenever belief in T violates Epistemic Parity. In the case of mereology, conservative theories subscribing to trees and fishes but not to trishes are arbitrary since they require us to believe in one theory in detriment of another epistemically on a par, namely, another conservative theory which also includes trishes. Generalising, any conservative theory will entail a similar violation of Epistemic Parity so that the only non-arbitrary mereological theory admitting cases of composition is mereological universalism. (Compare with Fairchild (2022, 625). To avoid mischaracterization let me stress that this argument is not endorsed there.)

Regardless of the previous argument's merits, the arbitrariness charge against strict finitism can be charitably understood in terms of the model of Epistemic Parity. The first thing to note

is that we are often at a loss about what could constitute even defeasible evidence that some specific $n \in \mathbb{N}$ is the largest number which was *not* at the same time evidence that $n + 1$ is the largest number. (Recall our convention of using '+' to refer to classical addition.) This last claim is easily misread, so to facilitate exposition define $\psi(x) := \forall y \exists z(x = y \oplus z)$. What I claim is that for any $n$, if $E$ is evidence that $\psi(\overline{n})$ is true in the strict finitist's correct model of arithmetical reality, then it is also evidence that $\psi(\overline{n+1})$ (or some other number) is true in the strict finitist's correct model of arithmetical reality. Note that this is not only due to familiar worries about our epistemic or causal access to abstract objects as per Benacerraf (1973). Even granting that we do possess some arithmetical knowledge, any of the usual reasons the strict finitist can give for $\psi(\overline{n})$ being true about arithmetical reality — e.g. $n$ has huge Kolmogorov complexity, or $n$ is larger than the number of particles in the universe, or it is 'so large that it has no physical or psychological significance' (Priest, 1994b, 8)[3] — are also reasons for $\psi(\overline{n+1})$ (or some other number) being true.[4] It follows that the theories $\psi(\overline{n})$ and $\psi(\overline{n+1})$ are epistemically on a par.[5]

Furthermore, once it is assumed that there is a largest number — that is, that $\exists x \psi(x)$ is true — it becomes natural to ask after the specific number which serves as a witness to the formula. There would be something epistemically disappointing were the strict finitist to hold that there is a largest number and *just leave it at that*, refusing to pinpoint the specific number which is in point of fact the largest. But we just saw that whatever answer the strict finitist gives in this instance — whatever the $\psi(\overline{n})$ to which the strict finitist ends up subscribing — is arguably underdetermined by the evidence. Therefore, any such answer will involve choosing among theories which are epistemically on a par.

Putting all these caveats together yields a new reading of the arbitrariness charge: strict finitism is arbitrary given that it violates Epistemic Parity — it requires us to believe that $\psi(\overline{n})$ in detriment of $\psi(\overline{n+1})$ (or some other number). Indeed, as I understand it, this is what is at

[3] I hasten to add that Priest's own position does not fall squarely within strict finitism.

[4] I ignore rather desperate replies like the idea that a god or some omniscient being could tell us whether $\psi(\overline{n})$ is the case or not. Those replies also apply to Benacerraf's argument and, much like in that case, they are beside the point — a philosophy of mathematics should not have to assume theism.

[5] By a 'theory' I just mean a collection of formulas, so that the single sentence $\psi(\overline{n})$ counts as a theory in my sense.

stake in Van Bendegem's own account of the arbitrariness charge that I quickly referenced at the start, namely:

> The essence of this objection is that the closer the strict finitist gets to a precise characterization of the greatest number, say $L$, the easier it becomes for adversaries to [...] to come up with questions about $L$ that make it seem quite arbitrary. [...] Classical examples of the latter are: Is $L$ prime? If not, what or how many are its prime factors? Can $L$ be divided by a specific number, say 3? (Van Bendegem, 2012, 144)

According to Van Bendegem, the strict finitist has to specify which number is the largest. However, in the process — 'the closer the strict finitist gets to a precise characterization of the greatest number' — she will end up subscribing to some theory $\psi(\overline{L})$. Assume for sake of argument that that $L$ is a prime number. Note that we are similarly at a loss about what could constitute even defeasible evidence that the largest number is prime. The theories 'the largest number is prime' and 'the largest number is not prime' are epistemically on a par. But by subscribing to $\psi(\overline{L})$, strict finitism seems to require us to believe that 'the largest number is prime' violating Epistemic Parity.

Is this a good objection to strict finitism?

### 5.4.2 Knowability: A Reply

To risk repetition, strict finitism posits the existence of a largest number. If the view is to have any intuitive appeal, it must add that it is knowable that this is so. The opposite claim in fact — that there is a largest number *and* also that it is not knowable that this is so — is reminiscent of a Moorean absurdity. And from this assumption, it is tempting to infer that for some $n \in \mathbb{N}$ it is knowable that $n$ is the largest number. Yet, this is the precise entailment which the strict finitist must guard herself against: this is because once it is assumed that for some $n$ it is knowable that $n$ is the largest number, it becomes natural to ask after the specific number which serves as a witness to the formula; and, as we saw, whatever answer the strict finitist gives here will

arguably involve choosing among theories which are epistemically on a par. On the contrary, I think that the proper reply to the epistemic version of the arbitrariness charge is to accept a certain amount of *epistemic* indeterminacy surrounding the largest number.

To explain. Before I had talked about the indeterminacy of a formula $\varphi$ in terms of there being no facts of the matter about whether $\varphi$. Still, it is not uncommon to talk about indeterminacy in another sense. Say that $\varphi$ is epistemically indeterminate if and only if it is unknowable whether $\varphi$ (see for example Field (2015, 155) for this different use). For purposes of exposition, I introduce some extra terminology. Let the formula '$\mathrm{K}_\Diamond\varphi$' abbreviate 'it is knowable that $\varphi$'. And let the formula '$\nabla\varphi$' abbreviate 'it is unknowable whether $\varphi$'. Given this we can say that $\varphi$ is epistemically indeterminate if and only if $\nabla\varphi$.[6] Finally, recall that $\psi(x) := \forall y \exists z(x = y \oplus z)$.

I have already suggested that the strict finitist should accept that it is knowable that there is a largest number, that is:

$$\mathrm{K}_\Diamond \ \ \exists x \psi(x) \qquad (\mathrm{K}_\Diamond \text{ Broad Scope})$$

But I propose further that, for *every* $n \in \mathbb{N}$, she reject the corresponding sentence:

$$\mathrm{K}_\Diamond \ \ \psi(\overline{n}) \qquad (\mathrm{K}_\Diamond \text{ Narrow Scope})$$

Moreover, in lieu of $\mathrm{K}_\Diamond$ Narrow Scope, the strict finitist should, for *some* $n \in \mathbb{N}$, accept the corresponding sentence:

$$\nabla \ \ \psi(\overline{n}) \qquad (\nabla \text{ Narrow Scope})$$

Essentially, the idea is that, according to strict finitism, for no number is it knowable that that number is the largest, even though there is some number for which it is unknowable. And since the strict finitist will most probably wish to claim that it is knowable that the number 2 is not

[6] A note on the terminology. The formula '$\mathrm{K}\varphi$' is usually reserved for (loosely) 'The agent knows that $\varphi$'. The subscript in '$\mathrm{K}_\Diamond$' serves as a reminder that 'knowability' is in part a modal concept — it involves the claim that something *can* be known. Even though for the sake of generality I am happy to treat the operator as a primitive, instead of reducing it to a possibility claim like $\Diamond\mathrm{K}\varphi$. Relatedly, we might want to define epistemic indeterminacy in terms of the knowability operator so that $\nabla\varphi := \neg(\mathrm{K}_\Diamond\varphi \vee \mathrm{K}_\Diamond\neg\varphi)$. But again for my purposes I can leave this issue open. Below I provide a clear semantics for the operators, making sense of how the strict finitist can decide which claims are knowable or epistemically indeterminate.

the largest number, we could add that $K_\Diamond \neg \psi(\overline{2})$ as well as other akin principles.

At this point recall that according to the epistemic reading of the arbitrariness charge, the strict finitist is forced to choose among theories which are epistemically on a par when asked after the specific number which is in point of fact the largest. As we saw, refusal to answer this question would be epistemically disappointing. But it seems that this is only so on the assumption that for some $n \in \mathbb{N}$ it is knowable that $n$ is the largest number. Once this is rejected, it is not clear why the strict finitist should ever feel the need to pinpoint the largest number. The opposite claim in fact — that I believe that $n$ is the largest number *and* that it is unknowable whether $n$ is the largest number — is again reminiscent of a Moorean absurdity. In sum: once $K_\Diamond$ Narrow Scope is dropped and some instance of $\nabla$ Narrow Scope stuck in its place, the strict finitist is free to suspend belief about the specific number which is the largest. In this case no violation of Epistemic Parity is in sight *pace* the arbitrariness charge.

### 5.4.3 A Model of Knowability within Strict Finitism

The previous subsection argued for epistemic indeterminacy with respect to the largest number as a way of responding to the arbitrariness charge. An immediate objection now states that the strict finitist's reasons for embracing epistemic indeterminacy are purely pragmatic: there is no justification for epistemic indeterminacy short of addressing the charge. But far from this being the case, the principles $K_\Diamond$ Broad Scope and $\nabla$ Narrow Scope are actually justified by certain intuitions to which the objector already subscribes. More precisely, the guiding intuition behind the epistemic version of the arbitrariness charge is that there does not seem to be any evidence for some $n \in \mathbb{N}$ being the largest number which is not at the same time evidence that $n + 1$ is the largest number. I think this is on the whole correct and that the strict finitist should agree. Nonetheless, I also think that if the strict finitist shares this intuition then she can motivate an account of knowability in arithmetic where the epistemic principles follow in a natural way.

For the strict finitist, arithmetic is correctly modelled by finite-domain structures as opposed to the classical models of first-order Peano Arithmetic (PA). One way to formalise this idea is by

reference to the theories developed by Van Bendegem (2003), building on the work of Mycielski (1981). Consider a first-order language $\mathcal{L} = \{0, S, \oplus, \otimes, c\}$ augmenting $\mathcal{L}_{\mathsf{PA}}$ with a constant $c$ whose intended meaning will be the largest number. Consider further the $\mathcal{L}$-theory $\mathsf{PA}^c$ whose non-logical axioms are: the first-order Induction Schema; the usual primitive recursive equations for addition and multiplication, that is, $\forall x(x \oplus 0 = x)$ and $\forall x \forall y(x \oplus S(y) = S(x \oplus y))$ and $\forall x(x \otimes 0 = 0)$ and $\forall x \forall y(x \otimes S(y) = (x \otimes y) \oplus x)$; the successor-axiom $\forall x(S(x) \neq 0)$ expressing that zero has no predecessors; and finally, the sentence $\forall x \forall y((x \neq c \wedge y \neq c) \rightarrow (S(x) = S(y) \rightarrow x = y))$. Unlike $\mathsf{PA}$ which contains a sentence expressing the injectivity of the successor function — namely, $\forall x \forall y(S(x) = S(y) \rightarrow x = y)$ — in the theory $\mathsf{PA}^c$, this sentence becomes 'relativized' to the largest number.

The crucial observation is that now, and unlike $\mathsf{PA}$, the theory $\mathsf{PA}^c$ is consistent with an extra formula of the form $S(c) = c$ stating that $c$ is the largest number. Semantically, this means that the theory has models whose successor-graph is just a copy of a finite initial segment of the classical models of arithmetic followed by a largest number equal to its own successor — in this model, the function 'loops' at the end.

By adding extra restrictions it is possible to control when the loop starts. In particular, for $n \in \mathbb{N}$, define the theory $\mathsf{PA}^c_{n+1}$ whose axioms are those of $\mathsf{PA}^c$ together with $S^{n+1}(0) = c$ and with $S(c) = c$. What do the models of $\mathsf{PA}^c_{n+1}$ look like? Well, we know that the axioms of $\mathsf{PA}^c$ guarantee that the models will start as a copy of an initial segment of the classical models; but also that the extra axioms guarantee that this segment ends at $n + 1$ — the largest number in the model — with the function looping after that. Since it will prove helpful later on, let me give a more detailed description of a model for $\mathsf{PA}^c_{n+1}$. The domain $D = \{0, ..., n + 1\} \subset \mathbb{N}$ will be a finite subset of the classical natural numbers, with $n+1$ being interpreted as the largest number in the model so that the interpretation of the new constant is $I(c) = n+1$. The successor function behaves 'as usual' up to the largest number — $S(m) = m+1$ for $m < n+1$ — whereas the largest number's successor is identified with itself — $S(n + 1) = n + 1$. For addition: if the classical sum of two numbers $m$ and $r$ is smaller or equal to the largest number, then $m \oplus r$ takes that classical value — i.e. $m \oplus r = p$ if $m + r = p \leq n + 1$ — otherwise, it is mapped to the

largest number itself — $m \oplus r = n + 1$. Similarly for multiplication $\otimes$. The successor-graph of a model of $\mathsf{PA}^c_{n+1}$ is depicted in Figure 5.1 (for vividness I assume that $n > 5$).[7]

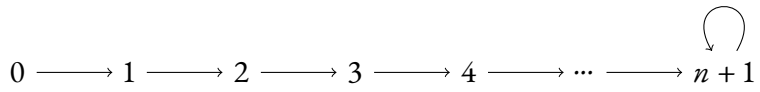



Figure 5.1: The successor-graph of a model of $\mathsf{PA}^c_{n+1}$.

For exposition, it is helpful to list together the axioms of $\mathsf{PA}$ and $\mathsf{PA}^c_{n+1}$ — I mark in bold where the axiomatizations diverge:

$\mathsf{PA}$

1. $\forall x(S(x) \neq 0)$
2. $\mathbf{\forall x \forall y\,(S(x) = S(y) \rightarrow x = y)}$
3. $\forall x(x \oplus 0 = x)$
4. $\forall x \forall y(x \oplus S(y) = S(x \oplus y))$
5. $\forall x(x \otimes 0 = 0)$
6. $\forall x \forall y(x \otimes S(y) = (x \otimes y) \oplus x)$
7. $(\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(S(x)))) \rightarrow \forall x \phi(x)$

$\mathsf{PA}^c_{n+1}$

1. $\forall x(S(x) \neq 0)$
2. $\mathbf{\forall x \forall y\,((x \neq c \wedge y \neq c) \rightarrow (S(x) = S(y) \rightarrow x = y))}$
3. $\forall x(x \oplus 0 = x)$
4. $\forall x \forall y(x \oplus S(y) = S(x \oplus y))$
5. $\forall x(x \otimes 0 = 0)$
6. $\forall x \forall y(x \otimes S(y) = (x \otimes y) \oplus x)$
7. $(\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(S(x)))) \rightarrow \forall x \phi(x)$
8. $\mathbf{S(c) = c}$
9. $\mathbf{S^{n+1}(0) = c}$

It can be shown that each theory $\mathsf{PA}^c_{n+1}$ is categorical (Van Bendegem, 2003, 238 Theorem 7) so that it provides a characterization of arithmetic where $n + 1$ is the largest number. Hence, we can formally regiment the thesis that a specific $n \in \mathbb{N}$ (such that $n \neq 0$) is a plausible candidate for the largest number by saying that the theory $\mathsf{PA}^c_n$ is a plausible characterization of arithmetic. Now, the strict finitist literature already recognises some plausible candidates for the largest number as in a number 'so large that it has no physical or psychological significance'

[7] It is interesting to compare this structure with the LP-models of arithmetic discussed in Chapter 2. See in particular Figure 2.2.

(Priest, 1994b, 8). For purposes of illustration let $n$ be one such candidate – in terms of our formal axiomatization, we are then assuming that for the strict finitist the theory $\mathsf{PA}^c_n$ is a plausible characterization of arithmetic: for her, it is an epistemic possibility that arithmetical reality is a model of $\mathsf{PA}^c_n$. Furthermore, recall that our strict finitist should agree with the rationale behind the epistemic reading of the arbitrariness charge, namely, that it would be arbitrary to suppose that *only* $n$ is a good candidate for the largest number, as opposed to also $n + 1$, or $n + 1 + 1$, ... (To repeat, there does not seem to be any evidence for $n$ being the largest number that is not evidence for $n + 1$.) As a result, the strict finitist should add that, for any $m \in \mathbb{N}$, the theory $\mathsf{PA}^c_{n+m}$ is also a plausible characterization of arithmetic.

Given this, how should the strict finitist go about deciding which sentences are knowable and which are epistemically indeterminate? A natural idea (inspired by supervaluationism) is to let knowability and epistemic indeterminacy be fixed by the joint contribution of each theory $\mathsf{PA}^c_{n+m}$ to truth, that is: a sentence $\varphi \in \mathcal{L}_{\mathsf{PA}}$ is knowable if, for every $m \in \mathbb{N}$, the models for $\mathsf{PA}^c_{n+m}$ all satisfy $\varphi$. More schematically, we have the following template:

$$\text{For } \varphi \text{ a sentence in } \mathcal{L}_{\mathsf{PA}} \begin{cases} \mathrm{K}_\Diamond \varphi & \text{if } \mathsf{PA}^c_{n+m} \models \varphi \text{ for all } m \in \mathbb{N} \\ \mathrm{K}_\Diamond \neg \varphi & \text{if } \mathsf{PA}^c_{n+m} \not\models \varphi \text{ for all } m \in \mathbb{N} \\ \nabla \varphi & \text{otherwise} \end{cases}$$

It is now easy to see how our earlier epistemic principles follow. Take an arbitrary model of $\mathsf{PA}^c_{n+m}$ for arbitrary $m \in \mathbb{N}$. For any number $s$ in the model's domain there is an element $r$ such that $s \oplus r = n + m$ (namely, the number $r$ such that classically $s + r = n + m$). Hence, the model satisfies $\forall y \exists z(\overline{n+m} = y \oplus z)$ and $\exists x \forall y \exists z(x = y \oplus z)$. Since $\mathsf{PA}^c_{n+m}$ is categorical we have $\mathsf{PA}^c_{n+m} \models \exists x \forall y \exists z(x = y \oplus z)$. Since $m \in \mathbb{N}$ was arbitrary it follows: $\mathrm{K}_\Diamond \exists x \forall y \exists z(x = y \oplus z)$ which is equivalent to $\mathrm{K}_\Diamond \exists x \psi(x)$. This expresses the claim that it is knowable that there is a largest number as per $\mathrm{K}_\Diamond$ Broad Scope. On the other hand, the element which plays the role of the largest number will change across different theories: indeed, for arbitrary $m \in \mathbb{N}$, whenever $\mathsf{PA}^c_{n+m} \models \forall y \exists z(\overline{n+m} = y \oplus z)$ it follows that $\mathsf{PA}^c_{n+m+1} \not\models \forall y \exists z(\overline{n+m} = y \oplus z)$. From this,

the strict finitist can conclude that $\nabla\forall y\exists z(\overline{n+m} = y \oplus z)$ as per $\nabla$ Narrow Scope. Finally, if there is some true instance of $K_\Diamond$ Narrow Scope, then there is some $s$ such that $\forall y\exists z(\bar{s} = y \oplus z)$ is entailed by *every* theory $\mathsf{PA}^c_{n+m}$. But the previous argument already shows that this cannot be the case. Hence, the strict finitist is right in rejecting this principle.

The preceding shows how it is possible to formulate an account of strict finitism where the relevant epistemic principles follow in a natural way. For these reasons, by accepting some amount of epistemic indeterminacy surrounding the largest number, I think that the epistemic reading of the arbitrariness charge can be resisted.

## 5.5 Conclusion

We are uneasy with the idea of the natural numbers suddenly stopping at a specific point along the natural line. Pre-theoretically at least, it looks as if the stopping point is bound to be unacceptably arbitrary. Nonetheless, not only is this arbitrariness charge subject to refinement in various ways, once so refined, it is clear that it can be resisted. I conclude that *if* it is to be shown that strict finitism is arbitrary, the objector will have to go beyond facile images about 'the natural numbers suddenly stopping' and be much clearer about the exact sense in which the view is said to be arbitrary.

# Conclusion

This thesis studied the problem of Arithmetical Determinacy. Chapter 1 argued that there is no unique way of understanding the concept of determinacy. Chapters 2 and 3 argued that popular computability constraints are unsuccessful in securing arithmetical determinacy. Chapter 4 motivated a new account of determinacy in terms of sentences possessing non-classical truth-values. Chapter 5 defended strict finitism, framing some objections against the view in terms of the concept of arithmetical determinacy.

# References


Assadian, B. (2018). The semantic plights of the ante-rem structuralist. *Philosophical studies*, *175*(12), 3195–3237.

Baker, A. (2007). Is There a Problem of Induction for Mathematics? In M. Leng, A. Paseau, & M. D. Potter (Eds.), *Mathematical Knowledge* (pp. 59–73). Oxford University Press.

Balaguer, M. (1995). A Platonist Epistemology. *Synthese*, *103*(3), 303–325.

Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics*. Oxford University Press.

Balaguer, M. (2001). A Theory of Mathematical Correctness and Mathematical Truth. *Pacific Philosophical Quarterly*, *82*(2), 87–114.

Barnes, E., & Williams, J. R. G. (2011). A Theory of Metaphysical Indeterminacy. In K. Bennett & D. W. Zimmerman (Eds.), *Oxford Studies in Metaphysics* (Vol. 6, pp. 103–148). Oxford University Press.

Beall, J. (2011). Multiple-Conclusion LP And Default Classicality. *The Review of Symbolic Logic*, *4*(2), 326–336.

Beall, J., Forster, T., & Seligman, J. (2013). A Note on Freedom from Detachment in the Logic of Paradox. *Notre Dame Journal of Formal Logic*, *54*(1), 15–20.

Belnap, N. D. (1977). A Useful Four-Valued Logic. In J. M. Dunn & G. Epstein (Eds.), *Modern Uses of Multiple-Valued Logic* (pp. 8–37). Reidel.

Benacerraf, P. (1962). Tasks, Super-Tasks and the modern Eleatics. *Journal of Philosophy*, *59*(24), 765–784.

Benacerraf, P. (1965). What Numbers Could Not Be. *Philosophical Review*, *74*(1), 47–73.

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

Benacerraf, P. (1996). Recantation or Any Old $\omega$-Sequence Would Do After All. *Philosophia Mathematica*, *4*(2), 184–189.

Benacerraf, P., & Putnam, H. ([1964] 1983). Introduction. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of Mathematics: Selected Readings* (2nd ed., pp. 1–38). Cambridge University Press.

Bennett, K. (2015). Perfectly Understood, Unproblematic, and Certain: Lewis on Mereology. In B. Loewer & J. Schaffer (Eds.), *A Companion to David Lewis* (pp. 250–261). Wiley-Blackwell.

Berry, S. (2021). Physical Possibility and Determinate Number Theory. *Philosophia Mathematica*, *29*(3), 299–317.

Berry, S. (2022). *A Logical Foundation for Potentialist Set Theory*. Cambridge University Press.

Berry, S. (2023). $\Sigma^0_1$ Soundness Isn't Enough: Number-theoretic Indeterminacy's Unsavory Physical Commitments. *The British Journal for the Philosophy of Science*, *74*(2), 469–484.

Berto, F. (2007). *How to Sell a Contradiction: The Logic and Metaphysics of Inconsistency*. College Publications.

Bigelow, J. (1988). *The Reality of Numbers: A Physicalist's Philosophy of Mathematics*. Oxford University Press.

Breckenridge, W., & Magidor, O. (2012). Arbitrary reference. *Philosophical Studies*, *158*(3), 377–400.

Bricker, P. (2016). Ontological Commitment. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/ontological-commitment/.

Brown, B. (2002). On Paraconsistency. In D. Jacquette (Ed.), *A Companion to Philosophical Logic* (pp. 628–650). Wiley-Blackwell.

Button, T. (2009). SAD Computers and Two Versions of the Church-Turing Thesis. *British Journal for the Philosophy of Science*, *60*(4), 765–792.

Button, T., & Smith, P. (2012). The Philosophical Significance of Tennenbaum's Theorem. *Philosophia Mathematica*, *20*(1), 114–121.

Button, T., & Walsh, S. (2018). *Philosophy and Model Theory*. Oxford University Press.

Cheng, Y. (2021). Current Research on Gödel's Incompleteness Theorems. *The Bulletin of Symbolic Logic*, *27*(2), 113–167.

Chihara, C. S. (1965). On the Possibility of Completing an Infinite Process. *Philosophical Review*, *74*(1), 74–87.

Chisholm, R. (1967). He Could Have Done Otherwise. *Journal of Philosophy*, *64*(13), 409–417.

Clarke-Doane, J. (2008). Multiple Reductions Revisited. *Philosophia Mathematica*, *16*(2), 244–255.

Clarke-Doane, J. (2020a). *Morality and Mathematics*. Oxford University Press.

Clarke-Doane, J. (2020b). Set-Theoretic Pluralism and the Benacerraf Problem. *Philosophical Studies*, *177*(7), 2013–2030.

Cowling, S. (2015). Non-Qualitative Properties. *Erkenntnis*, *80*(2), 275–301.

Cowling, S. (2016). Haecceitism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2016 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2016/entries/haecceitism/.

Cutland, N. (1980). *Computability: An Introduction to Recursive Function Theory*. Cambridge University Press.

Dean, W. (2014). Models and Computability. *Philosophia Mathematica*, *22*(2), 143–166.

Denyer, N. (1995). Priest's Paraconsistent Arithmetic. *Mind*, *104*(415), 566–575.

Dodd, J. (2002). Is Truth Supervenient on Being? *Proceedings of the Aristotelian Society*, *102*, 69–85.

Dummett, M. ([1973] 1978). The Philosophical Basis of Intuitionistic Logic. In H. E. R. . J. C. Shepherdson (Ed.), *Logic colloquium '73*. North-Holland. (Reprinted in Dummett (1978), pages 215–247)

Dummett, M. (1978). *Truth and Other Enigmas*. Duckworth.

Earman, J., & Norton, J. D. (1993). Forever is a Day: Supertasks in Pitowsky and Malament-

Hogarth Spacetimes. *Philosophy of Science*, *60*(1), 22–42.

Earman, J., & Norton, J. D. (1996). Infinite Pains: The Trouble with Supertasks. In A. Morton & S. P. Stich (Eds.), *Benacerraf and His Critics* (pp. 231–259). Blackwell.

Etesi, G., & Németi, I. (2002). Non-Turing Computations Via Malament-Hogarth Space-Times. *International Journal of Theoretical Physics*, *41*, 341–70.

Fairchild, M. (2022). Arbitrariness and the Long Road to Permissivism. *Noûs*, *56*(3), 619–638.

Feferman, S. (1998). *In the Light of Logic*. Oxford University Press.

Feferman, S. (2011). *Is the Continuum Hypothesis a definite mathematical problem?* Retrieved from http://logic.harvard.edu/EFI_Feferman_IsCHdefinite.pdf (Manuscript for: *Exploring the Frontiers of Incompleteness, Harvard*. Accessed on 23-06-2021)

Feferman, S. (2013). *Comments for EFI Workshop Meet.* (Manuscript for: *Exploring the Frontiers of Incompleteness, Harvard*)

Feferman, S., Friedman, H. M., Maddy, P., & Steel, J. R. (2000). Does Mathematics Need New Axioms? *The Bulletin of Symbolic Logic*, *6*(4), 401–446.

Ferguson, T. M. (2019). Variations on the Collapsing Lemma. In C. Başkent & T. M. Ferguson (Eds.), *Graham Priest on Dialetheism and Paraconsistency* (pp. 249–270).

Field, H. ([1998] 2001a). Mathematical Objectivity and Mathematical Objects. In S. L. C. MacDonald (Ed.), *Contemporary Readings in the Foundations of Metaphysics* (pp. 387–403). Blackwell. (Reprinted in Field (2001), pages 315–331)

Field, H. ([1998] 2001b). Which Undecidable Mathematical Sentences Have Determinate Truth Values? In H. G. Dales & G. Oliveri (Eds.), *Truth in Mathematics* (pp. 291–310). Oxford University Press. (Reprinted in Field (2001), pages 332–360)

Field, H. (2001). *Truth and the Absence of Fact*. Oxford University Press.

Field, H. (2015). Mathematical Undecidables, Metaphysical Realism, and Equivalent Descriptions. In R. E. Auxier, D. R. Anderson, & L. E. Hahn (Eds.), *The Philosophy of Hilary Putnam* (Vol. XXXIV, pp. 145–172). Open Court.

Field, H. (2016). *Science Without Numbers: A Defence of Nominalism* (2nd ed.). Oxford

University Press.

Fine, K. (1975). Vagueness, Truth and Logic. *Synthese*, *30*(3–4), 265–300.

Fine, K. (2005). The Varieties of Necessity. In *Modality and Tense: Philosophical Papers* (pp. 235–260). Oxford University Press.

Florio, S., & Linnebo, O. (2016). On the Innocence and Determinacy of Plural Quantification. *Noûs*, *50*(3), 565–583.

Florio, S., & Linnebo, O. (2021). *The Many and the One: A Philosophical Study of Plural Logic*. Oxford: Oxford University Press.

Frixione, M. (2001). Tractable Competence. *Minds and Machines*, *11*, 379–397.

Gaifman, H. (2003). Non-Standard Models in a Broader Perspective. In A. Enayat & R. Kossak (Eds.), *Nonstandard Models of Arithmetic and Set Theory: AMS Special Session Nonstandard Models of Arithmetic and Set Theory, January 15-16, 2003, Baltimore, Maryland* (Vol. 361, pp. 1–22). American Mathematical Society.

George, A. (1993). How Not to Refute Realism. *The Journal of Philosophy*, *90*(2), 53-72.

Goodship, L. (1996). On Dialethism. *Australasian Journal of Philosophy*, *74*(1), 153–161.

Halbach, V., & Horsten, L. (2005). Computational Structuralism. *Philosophia Mathematica*, *13*(2), 174–186.

Hale, B. (2020). The Logic of Metaphysical Modality. In O. Bueno & S. A. Shalkowski (Eds.), *The Routledge Handbook of Modality* (pp. 308–318). Routledge.

Hamkins, J. D. (2012). The Set-Theoretic Multiverse. *Review of Symbolic Logic*, *5*(3), 416–449.

Hamkins, J. D. (2015). Is the Dream Solution of the Continuum Hypothesis Attainable? *Notre Dame Journal of Formal Logic*, *56*(1), 135–145.

Hamkins, J. D. (2018). *The modal logic of arithmetic potentialism and the universal algorithm.* arXiv. Retrieved from https://arxiv.org/abs/1801.04599

Hamkins, J. D., & Yang, R. (2013). *Satisfaction is not absolute.* arXiv. Retrieved from https://arxiv.org/abs/1312.0670

Hart, W. D. (1996). *The Philosophy of Mathematics*. Oxford University Press.

Hawthorne, J. (2006). *Metaphysical Essays*. Clarendon Press.

Hellman, G. (1989). *Mathematics Without Numbers: Towards a Modal-Structural Interpretation*. Oxford University Press.

Hellman, G. (2001). Three Varieties of Mathematical Structuralism. *Philosophia Mathematica*, *9*(2), 184–211.

Hellman, G. (2005). Structuralism. In S. Shapiro (Ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic* (pp. 536–562). Oxford University Press.

Hogarth, M. (1992). Does General Relativity Allow an Observer to View an Eternity in a Finite Time? *Foundations Of Physics Letters*, *5*, 173–181.

Hogarth, M. (1994). Non-Turing Computers and Non-Turing Computability. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1994*, 126–138.

Hogarth, M. (2004). Deciding Arithmetic Using SAD Computers. *British Journal for the Philosophy of Science*, *55*(4), 681–691.

Hogarth, M. (2009). Non-Turing Computers are the New Non-Euclidean Geometries. *International Journal of Unconventional Computing*, *5*, 277–291.

Horsten, L. (2001). On our ability to fix intended structures. In A. A. Derksen (Ed.), *Moving Ahead: Philosophy of Mind and Realism* (pp. 121–134). Dutch University Press.

Horsten, L. (2012). Vom Zahlen zu den Zahlen: On the Relation Between Computation and Arithmetical Structuralism. *Philosophia mathematica*, *20*(3), 275–288.

Hyde, D. (1997). From Heaps and Gaps to Heaps of Gluts. *Mind*, *106*(424), 641–660.

Hyde, D. (2010). The Prospects of a Paraconsistent Response to Vagueness. In R. Dietz & S. Moruzzi (Eds.), *Cuts and Clouds: Vaguenesss, its Nature and its Logic* (pp. 385–405). Oxford University Press.

Hyde, D., & Raffman, D. (2018). Sorites Paradox. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2018 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/sorites-paradox/.

Isaacson, D. ([1987] 1996). Arithmetical Truth and Hidden Higher-Order Concepts. In *Studies in Logic and the Foundations of Mathematics* (Vol. 122, pp. 147–169). Elsevier. (Reprinted

in Hart (1996), pages)

Isaacson, D. (2011). Necessary and sufficient conditions for undecidabillity of the Gödel sentence and its truth. In P. Clark, D. DeVidi, & M. Hallett (Eds.), *Vintage Enthusiasms: Essays in Honour of John Bell* (pp. 135–152). Heidelberg and New York: Springer Verlag.

Jaśkowski, S. ([1948] 1969). Rachunek zdań dla systemów dedukcyjnych sprzecznych. *Studia Societatis Scientiarum Torunensi (Sectio A)*, *1*(5), 55–77. (English translation appeared as (1969). Propositional Calculus for Contradictory Deductive Systems. *Studia Logica*, 24, 143—157)

Kaye, R. (1991). *Models of Peano Arithmetic* (Vol. 15). Clarendon Press.

Kaye, R. (2011). Tennenbaum's Theorem for Models of Arithmetic. In J. Kennedy & R. Kossak (Eds.), *Set Theory, Arithmetic, and Foundations of Mathematics: Theorems, Philosophies* (pp. 66–79). Cambridge University Press.

Keefe, R. (2000). *Theories of Vagueness*. Cambridge University Press.

Korman, D. Z. (2010). Strange Kinds, Familiar Kinds, and the Charge of Arbitrariness. In D. W. Zimmerman (Ed.), *Oxford Studies in Metaphysics* (pp. 119–144). Oxford University Press.

Korman, D. Z. (2020). Ordinary Objects. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/ordinary-objects/.

Kotlarski, H. (1991). Full Satisfaction Classes: A Survey. *Notre Dame Journal of Formal Logic*, *32*(4), 573–579.

Kotlarski, H. (2019). *A Model-Theoretic Approach to Proof Theory* (Vol. 51; Z. Adamowicz, T. Bigorajska, & K. Zdanowski, Eds.). Springer.

Kotlarski, H., Krajewski, S., & Lachlan, A. H. (1981). Construction of Satisfaction Classes for Nonstandard Models. *Canadian Mathematical Bulletin*, *24*(3), 283–293.

Krajewski, S. (1974). Mutually Inconsistent Satisfaction Classes. *Bulletin de l'Académie Polonaise des Sciences: Série de sciences math., astr. et phys.*, *22*(10), 983–987.

Krajewski, S. (1976). Non-standard Satisfaction Classes. In W. Marek, M. Srebrny, & A. Zarach

(Eds.), *Set Theory and Hierarchy Theory: A Memorial Tribute to Andrzej Mostowski* (Vol. 537, pp. 121–144). Springer Verlag.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.

Lachlan, A. H. (1981). Full Satisfaction Classes and Recursive Saturation. *Canadian Mathematical Bulletin*, *24*(3), 295–297.

Lewis, D. K. (1986). *On the Plurality of Worlds*. Wiley-Blackwell.

Lewis, D. K. (2001). Truthmaking and Difference-Making. *Noûs*, *35*(4), 602–615.

Liggins, D. (2005). Truthmakers and Explanation. In H. Beebee & J. Dodd (Eds.), *Truthmakers: The Contemporary Debate* (pp. 105–115). Clarendon Press.

Linnebo, Ø., & Shapiro, S. (2019). Actual and Potential Infinity. *Noûs*, *53*(1), 160–191.

Lowe, E. J. (1987). Not a Counterexample to Modus Ponens. *Analysis*, *47*(1), 44–47.

Maddy, P. (2011). *Defending the Axioms: On the Philosophical Foundations of Set Theory*. Oxford University Press.

Manchak, J., & Roberts, B. W. (2022). Supertasks. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2022/entries/spacetime-supertasks/.

Markosian, N. (1998). Brutal composition. *Philosophical Studies*, *92*(3), 211–249.

Martin, D. A. (2018). *Completeness or Incompleteness of Basic Mathematical Concepts.* Retrieved from https://www.math.ucla.edu/~dam/booketc/efi.pdf (Accessed on 23-06-2021)

McGee, V. (1985). A Counterexample to Modus Ponens. *The Journal of Philosophy*, *82*(9), 462–471.

McGee, V. (1997). How We Learn Mathematical Language. *The Philosophical Review*, *106*(1), 35–68.

McGee, V. (2005). Afterword: Trying (with Limited Success) to Demarcate the Disquotational-Correspondence Distinction. In B. P. Amour-Garb & J. C. Beall (Eds.), *Deflationary Truth* (pp. 143–152). Open Court.

McSweeney, M. M. (2019). Following logical realism where it leads. *Philosophical studies*, *176*(1), 117–139.

Merricks, T. (2007). *Truth and Ontology*. Clarendon Press.

Murawski, R. (1997). Satisfaction Classes — a Survey. In R. Murawski & J. Pogonowski (Eds.), *Euphony and Logos: Essays in Honour of Maria Steffen-Batóg and Tadeusz Batóg* (Vol. 57, pp. 259–282). Brill.

Mycielski, J. (1981). Analysis Without Actual Infinity. *The Journal of Symbolic Logic*, *46*(3), 625–633.

Omori, H. (2016). From Paraconsistent Logic to Dialetheic Logic. In P. Verdée & H. Andreas (Eds.), *Logical Studies of Paraconsistent Reasoning in Science and Mathematics* (p. 111–134). Cham, Switzerland: Springer Verlag.

Parsons, C. (1990). The Uniqueness of the Natural Numbers. *Yyun: The Jerusalem Philosophical Quarterly*, *39*, 13–44.

Parsons, C. (2007). Mathematical Induction. In *Mathematical Thought and its Objects* (pp. 264–315). Cambridge University Press.

Parsons, J. (1999). There is No 'Truthmaker' Argument Against Nominalism. *Australasian Journal of Philosophy*, *77*(3), 325–334.

Picollo, L., & Waxman, D. (in press). Internalism and the Determinacy of Mathematics. *Mind*.

Piller, C. (1996). Vann McGee's Counterexample to Modus Ponens. *Philosophical Studies*, *82*(1), 27–54.

Priest, G. (1994a). Is Arithmetic Consistent? *Mind*, *103*(411), 337–349.

Priest, G. (1994b). What Could the Least Inconsistent Number Be? *Logique et Analyse*, *37*(145), 3–12.

Priest, G. (1996). On Inconsistent Arithmetics: A Reply to Denyer. *Mind*, *105*(420), 649–659.

Priest, G. (1997). Inconsistent Models of Arithmetic Part I: Finite Models. *Journal of Philosophical Logic*, *26*(2), 223–235.

Priest, G. (2006). *In Contradiction: A Study of the Transconsistent* (Expanded ed.). Clarendon Press.

Priest, G. (2017). What If? The Exploration of an Idea. *Australasian Journal of Logic*, *14*(1), 54–127.

Priest, G., & Routley, R. (1989). Systems of Paraconsistent Logic. In G. Priest, R. Routley, & J. Norman (Eds.), *Paraconsistent Logic: Essays on the Inconsistent* (pp. 142–155).

Putnam, H. (1967). Mathematics without Foundations. *The Journal of philosophy*, *64*(1), 5–22.

Putnam, H. (1977). Realism and Reason. *Proceedings and Addresses of the American Philosophical Association*, *50*(6), 483–498.

Putnam, H. (1980). Models and Reality. *Journal of Symbolic Logic*, *45*(3), 464–482.

Putnam, H. (1981). *Reason, Truth and History*. Cambridge University Press.

Putnam, H. (2015). Reply to Hartry Field. In R. E. Auxier, D. R. Anderson, & L. E. Hahn (Eds.), *The Philosophy of Hilary Putnam* (Vol. XXXIV, pp. 173–179). Open Court.

Quinon, P., & Zdanowski, K. (2011). *The Intended Model of Arithmetic. An Argument from Tennenbaum's Theorem.* Retrieved from https://www.impan.pl/~kz/files/PQKZ_Tenn.pdf (Accessed on 29-06-2023)

Relaford-Doyle, J., & Núñez, R. (2018). Beyond Peano: Looking Into the Unnaturalness of Natural Numbers. In S. Bangu (Ed.), *Naturalizing Logico-Mathematical Knowledge: Approaches From Philosophy, Psychology, and Cognitive Science* (pp. 234–251). Routledge.

Rosser, B. (1936). Extensions of Some Theorems of Gödel and Church. *Journal of Symbolic Logic*, *1*(3), 87–91.

Rosser, B. (1937). Gödel Theorems for Non-Constructive Logics. *Journal of Symbolic Logic*, *2*(3), 129–137.

Russell, B. ([1907] 1973). The Regressive Method for Discovering the Premises of Mathematics. In D. Lackey (Ed.), *Essays in Analysis* (pp. 272–283). London: George Allen & Unwin Ltd.

Russell, B. (1936). VII.—The Limits of Empiricism. *Proceedings of the Aristotelian Society*, *36*(1), 131-150.

Sayward, C. (1990). Four Views of Arithmetical Truth. *The Philosophical Quarterly*, *40*(159), 155–168.

Shapiro, S. (1991). *Foundations Without Foundationalism: A Case for Second-Order Logic*. Oxford University Press.

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

Shapiro, S. (2006). Structure and Identity. In F. MacBride (Ed.), *Identity and Modality* (pp. 109–145). Oxford University Press.

Shoemaker, S. (1998). Causal and Metaphysical Necessity. *Pacific Philosophical Quarterly*, *79*(1), 59–77.

Sidelle, A. (2002). On the Metaphysical Contingency of Laws of Nature. In J. Hawthorne & T. S. Gendler (Eds.), *Conceivability and Possibility* (pp. 309–336). Oxford University Press.

Sider, T. (2007). Neo-Fregeanism and Quantifier Variance. *Aristotelian Society Supplementary Volume*, *81*(1), 201–232.

Sinnott-Armstrong, W., Moor, J., & Fogelin, R. (1986). A Defense of Modus Ponens. *The Journal of Philosophy*, *83*(5), 296–300.

Smith, D. C. (2003). What is so Magical About a Theory of Intrinsic Intensionality? *Philosophical Papers*, *32*(1), 83–95.

Smith, P. (2013). *An Introduction to Gödel's Theorems* (2nd ed.). Cambridge University Press.

Smith, S. T. (1987). Nonstandard Characterizations of Recursive Saturation and Resplendency. *The Journal of Symbolic Logic*, *52*(3), 842–863.

Smith, S. T. (1989). Nonstandard Definability. *Annals of Pure and Applied Logic*, *42*(1), 21–43.

Sosa, E. (1999). Existential Relativity. *Midwest Studies in Philosophy*, *23*(1), 132–143.

Suszko, R. (1975). Remarks on Lukasiewicz's Three-Valued Logic. *Bulletin of the Section of Logic*, *4*(3), 87–90.

Swoyer, C. (1982). The Nature of Natural Laws. *Australasian Journal of Philosophy*, *60*(3), 203–233.

Thomson, J. (1954). Tasks and Supertasks. *Analysis*, *15*(1), 1–13.

Thomson, J. (1970). Comments on Professor Benacerraf's Paper. In W. C. Salmon (Ed.), *Zeno's*

*Paradoxes* (pp. 130–138). Bobbs-Merrill.

Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230-265.

Väänänen, J. (2021). Tracing Internal Categoricity. *Theoria*, *87*(4), 986–1000.

Väänänen, J., & Wang, T. (2015). Internal Categoricity in Arithmetic and Set Theory. *Notre Dame Journal of Formal Logic*, *56*(1), 121–134.

Van Bendegem, J. P. (2003). Classical arithmetic is quite unnatural. *Logic and Logical Philosophy*(11–12), 231–249.

Van Bendegem, J. P. (2012). A Defense of Strict Finitism. *Constructivist Foundations*, *7*(2), 141–149.

Van Cleve, J. (1986). Mereological Essentialism, Mereological Conjunctivism, and Identity Through Time. *Midwest Studies in Philosophy*, *11*(1), 141–156.

Van Rooij, I. (2008). The Tractable Cognition Thesis. *Cognitive Science*, *32*(6), 939-984.

Warren, J. (2020a). Killing Kripkenstein's Monster. *Noûs*, *54*(2), 257–289.

Warren, J. (2020b). *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. Oxford University Press.

Warren, J., & Waxman, D. (2020a). A Metasemantic Challenge for Mathematical Determinacy. *Synthese*, *197*(2), 477–495.

Warren, J., & Waxman, D. (2020b). Supertasks and Arithmetical Truth. *Philosophical Studies*, *177*(5), 1275–1282.

Welch, P. D. (2008). The Extent of Computation in Malament-Hogarth Spacetimes. *British Journal for the Philosophy of Science*, *59*(4), 659–674.

Weston, T. (1976). Kreisel, the Continuum Hypothesis and Second Order Set Theory. *Journal of Philosophical Logic*, *5*(2), 281–298.

Wetzel, L. (1989). That Numbers Could Be Objects. *Philosophical Studies*, *56*(3), 273–92.

Williamson, T. (1994). *Vagueness*. Routledge.

Woodin, W. H. (2011). A Potential Subtlety Concerning the Distinction between Determinism and Nondeterminism. In M. Heller & W. H. Woodin (Eds.), *Infinity: New Research*

*Frontiers* (pp. 119–129).

Yablo, S. (1987). Identity, Essence, and Indiscernibility. *Journal of Philosophy*, *84*(6), 293–314.

Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research*, *53*(1), 1–42.