Carnap and the Ontology of Mathematics

Benjamin Marschall
Darwin College, University of Cambridge

February 2020
This thesis is submitted for the degree of Doctor of Philosophy.

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Abstract

Carnap and the Ontology of Mathematics
Benjamin Marschall

In this thesis I investigate Rudolf Carnap's philosophy of mathematics. Most philosophers assume that the nature of mathematics raises deep philosophical questions, which call for theories about how we manage to know about and interact with abstract objects. Carnap's position, in contrast, is deflationary: he aims to show that we can take mathematics at face value *without* having to answer questions about the metaphysical status of mathematical objects. If Carnap is right, there is thus no need for a *philosophy* of mathematics as it is usually understood at all. The main argument of my thesis is that Carnap's position is unstable, since his own commitments force him to make at least *some* ontological assumptions about syntax, i.e. entities such as letters, strings, and proofs.

My claim that Carnap needs to accept some ontological questions as being in good shape goes against the received view in the secondary literature. Since the late 1980s interest in Carnap's philosophy of mathematics has been growing, mostly in the wake of important papers by Michael Friedman, Warren Goldfarb, and Thomas Ricketts. These and other scholars have forcefully defended Carnap against objections by, among others, Kurt Gödel, W. V. Quine, and Hilary Putnam. The most powerful challenge to Carnap's view, however, can actually be found in a less well-known paper by the logician E. W. Beth. The core of my thesis is thus a new interpretation of what I call *Beth's argument from non-standard models*, which relies on Gödel's incompleteness theorems and targets Carnap's claim that mathematics is analytic. I show that my reconstruction of Beth's argument is more charitable to the text than competing interpretations in the secondary literature, and argue that it is also more powerful since extant defences of Carnap cannot be applied.

Acknowledgements

I have been doing philosophy for nearly ten years now, and there are many people to thank. Starting with those who most directly contributed to making this thesis a reality, I thank Tim Button for being an amazing supervisor: always positive and encouraging, always full of good ideas, and as engaged as ever even after leaving Cambridge for UCL. I could not have written this thesis without him.

Alex Oliver was an excellent secondary supervisor, especially during the final stretch when he read and commented on a full draft. I know that not everyone is in the lucky position to have both their supervisors read their thesis before submission, so I am very grateful for the support. I also thank Michael Potter who, in a memorable first-year review meeting together with Alex, suggested that looking at the history of analytic philosophy might be the best way to tackle the issues I was interested in. I had considered working on Carnap before, but was put off by the fact that this seemed hard work. I am glad that this time I received a push in the right direction at the right moment.

The philosophy graduate community at Cambridge has made the time during which I have been working on this thesis particularly enjoyable. Special thanks go to Annie Bosse, Bertie Probyn, Lucy McDonald, Nathan Hawkins, Makan Nojoumian, Paula Keller, Senthuran Bhuvanendra, Sofía Melendez Gutierrez, Wouter Cohen, and the honorary graduate students Alex Jackson and Thomas Schindler. Many of them have given helpful feedback and advice more than once, and I hope that all at least somewhat enjoyed my Carnap-anecdotes.

The late Hugh Mellor not only was an unusually warm and welcoming presence in the Faculty, but, to my amazement, even gave me a copy of the Schilpp-volume on Carnap with scribbled notes by Paul Feyerabend in it. He is greatly missed.

In early 2020 I visited the University of Pittsburgh for a term in order to work with Thomas Ricketts. I thank Tom for being so generous with his time in order to talk through Beth and Carnap, for yet another book present (this time a copy of *The Logical Syntax of Language*), and for being extremely supportive during a turbulent mid-March. The meetings with him were very energising and led me to re-think my interpretation of Beth, which ultimately enabled me to give a much clearer account of the core argument. Consulting the unpublished material in the Rudolf Carnap Papers was also stimulating, and I thank the always helpful staff at Archives of Scientific Philosophy. Other people who made my time in Pittsburgh pleasant were Christian Feldbacher-Escamilla, Clemens Wetcholowsky, Patrick Chandler, Sander Verhaegh, Sophia Arbeiter, and Stephen Mackereth.

Before coming to Cambridge I studied in Berlin and St Andrews. For Berlin I first and foremost thank Tobias Rosefeldt for the continued support, and also Barbara Vetter, Emanuel Viebahn, Lukas Lewerentz, Karl-Georg Niebergall, Martin Aigner, Olaf Müller, Roland Krause, and Timo Schobinger for being inspirational in various ways. For St Andrews greetings go to the Stirling Bus Crew, especially to Jakob Ohlhorst and Xintong Wei.

Material drawn from this thesis has been presented at talks in Bratislava, Cambridge, Munich, Oxford, and Tilburg, and I thank the respective audiences for their comments and questions. Papers based on the thesis are or have been under review at *Erkenntnis* and the *Canadian Journal of Philosophy*, and many of the referee reports I received have been helpful and constructive. I also thank Eckehart Köhler for a long conversation about the relationship between Carnap and Gödel, Gary Ebbs for giving detailed answers to my questions about his interpretation of Carnap via email, Hannes Leitgeb for sending me a version of his and André Carus' *SEP* entry on Carnap before it went online, and Jared Warren for sending me a pre-print of his intriguing book-length defence of mathematical conventionalism.

The Arts and Humanities Research Council, the Cambridge Trust, the Aristotelian Society, the Royal Institute of Philosophy, and the German-American Fulbright Commission provided the funding that made my research possible in the first place, and for this I am very grateful.

It is unhealthy to think about philosophy all the time, and so thanks go to the members of the Cambridge University Rambling Club for igniting my passion for the East Anglian countryside.

I thank my parents for always being supportive and enthusiastic about whatever I do, even though I often struggle to explain the exact nature of my pursuits to them. Last but certainly not least, I thank Clara for everything.

Contents

Introduction

> *Gödel:* With indefinite concepts theology begins! *?!*
>
> He later says: he means thereby that in mutual communication one always needs to start from the definite. But of course he wants to use indefinite concepts too. (Carnap, note from July 1933, quoted from Köhler 2002: 123)

In this thesis I investigate Rudolf Carnap's philosophy of mathematics. Carnap claims that his account can vindicate classical mathematics, and hence makes revisionary projects like intuitionism or nominalism unnecessary, without committing him to any contentious metaphysical assumptions. My focus will be on the question whether Carnap's response to Gödel's incompleteness results is satisfactory, and I will argue for a negative answer. Carnap either needs to accept some *ontological* questions as meaningful and factual, or has to endorse a revisionary conception of mathematics after all.

One might wonder why an investigation as detailed as this one is needed in the first place. Of course to this day there is little consensus on which approach to the philosophy of mathematics is the most fruitful. One thing that does seem to be commonly accepted, however, is that the account Carnap develops in his *Logical Syntax of Language* (Carnap 1937a) is not a contender that needs to be taken very seriously. It would for instance be odd for someone to teach an introductory course into the philosophy of mathematics without mentioning Hilbert's formalism or Frege's logicism, but in my experience few people would expect extended discussions of Carnap's approach. The implicit assumption seems to be that Quine's arguments against logical conventionalism and analyticity (Quine 1949, Quine 1951) – and maybe also Gödel's incompleteness theorems – conclusively establish that Carnap's approach is fundamentally flawed.[1]

[1] Øystein Linnebo's introductory textbook in the philosophy of mathematics from 2017 does not mention Carnap once (Linnebo 2020a), for instance, and neither does Mark Colyvan's textbook from 2012 (Colyvan 2012). Things have not always been this way, however: the 1983 second edition of Paul Benacerraf and Hilary Putnam's influential selection of articles on

A very different picture emerges, however, if one spends some time studying the scholarly work that has been done on Carnap since the late 1980s, pioneered by Michael Friedman, Richard Creath, Warren Goldfarb and Thomas Ricketts, and continued by, among others, Gary Ebbs, Steve Awodey, André Carus, and Gregory Lavers.[2] By now there is a large number of articles which defend the coherence of Carnap's account of mathematics against numerous objections, including the most influential ones by Quine and Gödel. A keen reader of this literature will thus be tempted to ask why, given the apparent attractiveness of Carnap's view, not *everyone* is a Carnapian. In other words, the continued existence of foundational disputes about the nature of mathematics will seem puzzling if these scholars are on the right track.

My main argument goes against this latter-day trend, as I want to show that there *are* some serious problems with Carnap's position. This is not to say that the scholarly work I alluded to is misguided, however. As we will see throughout the thesis, Carnap can actually give convincing responses to a number of arguments commonly thought to be damaging, and hence those who argue that Carnap's position is defensible are largely on the right track. I will show, however, that there is a less well-known objection whose force has been underestimated by friends and enemies of Carnap so far: namely E. W. Beth's argument from non-standard models (Beth 1963). The core of the thesis is consequently a novel reconstruction of Beth's argument, to which, so I will argue, extant defences of Carnap do not apply.[3]

Methodologically speaking, the biggest challenge for my project is to present an argument that actually poses a problem for *Carnap*, and not merely a straw man who superficially resembles him. This worry is salient since the Carnap-scholars I mentioned have convincingly shown that many anti-Carnapian arguments miss their intended target, by ascribing views to Carnap he did not

the philosophy of mathematics still contains two paper's by Carnap (including "Empiricism, Semantics, and Ontology"), as well as Quine's paper "Carnap and Logical Truth" (Benacerraf and Putnam 1983). Stewart Shapiro also has a section on Carnap in his *Thinking about Mathematics* from 2000 (Shapiro 2000: section 5.3).

2 Some of the key works are Friedman 1999a, Creath 1990, Goldfarb and Ricketts 1992, Ricketts 1994, Friedman 1999b, Ebbs 1997, Awodey and Carus 2004, Lavers 2008, and I will discuss many other papers in what follows. Another important figure is Alan Richardson (Richardson 1994), whose work, however, tends to concentrate on Carnap's *Der Logische Aufbau der Welt* rather than his philosophy of logic and mathematics (Richardson 1997).

3 I am aware of two other PhD theses written on Carnap's philosophy of mathematics in the last ten years: Doyle 2013 and Friedman-Biglin 2015, which, broadly speaking, defend rather than attack Carnap's position. Friedman-Biglin develops a new version of the established responses to Gödel's arguments against Carnap, but does not discuss Beth. And while Doyle attacks the deflationary reading of Carnap associated with Goldfarb and Ricketts, he does not think that Beth's argument is a deep challenge.

himself hold. The consensus that has been emerging in the literature is that Carnap's overall position is a *deflationary* one, meaning that Carnap is not interested in even addressing many of the questions that other philosophers have meant to answer.

I will argue that there are good reasons to think that Beth's argument, as I interpret it, succeeds as an objection to Carnap's actual position. The exegetical situation is not straightforward, however. There certainly are passages which suggest that Carnap's deflationism is so radical that he does not have to worry about Beth's argument either. I will present some new and previously unpublished textual evidence that speaks against such an interpretation, however, and furthermore argue that a radically deflationary reading of Carnap is also hard to swallow on systematic grounds. This will probably not convince everyone, but I hope that even those who still want to side with Carnap will agree that, as is demonstrated by Beth, Carnap's approach of dealing with Gödel's incompleteness results requires far more attention than has been recognised so far.

I will briefly summarise the structure of the thesis, which is divided into three parts. The chapters in part I introduce Carnap's position in the philosophy of mathematics, which I call *internal Platonism*.[4] Chapter one outlines the crucial notion of a linguistic framework, and explains the role analyticity plays in the overall account. I also discuss to what extent Carnap's position is a form of linguistic conventionalism. In chapter two I describe Carnap's reaction to Gödel's incompleteness results, which involves using non-recursive rules to define 'analytic' for mathematical discourse. I then discuss an argument against Carnap's position put forward by Gödel himself, and show that it is uncertain how damaging Gödel's objection really is. The short chapter three concludes this part by briefly discussing Quine's famous rejection of the analytic/synthetic-distinction, as this will be helpful to appreciate how Beth's argument differs from more familiar concerns about analyticity.

In part II I then present my reading of Beth's *argument from non-standard models*, and argue that it is both more damaging to Carnap and more charitable to Beth than alternative interpretations found in the secondary literature. In chapter four I explain why, on my reading, Beth's argument can be used to criticise two of the main tenets of Carnap's account: the *principle of tolerance*, according to which choosing between different systems of logic is a purely pragmatic question, and the *internalist* thesis that one can quantify over abstract objects without any metaphysical commitments. The overall result is that these two theses need

4 This position is distinct from the internal platonism about set theory of Potter 2004: 40.

to be restricted, and thus Carnap is pushed into accepting at least *some* ontological questions as being meaningful and factual. Unfortunately Beth's paper is at times unclear and misleading, and it is therefore no wonder that commentators have interpreted his argument in various ways. In chapter five I survey the extant secondary literature and argue that my interpretation is preferable overall. In chapter six I respond to the worry that Beth merely attacks a straw man rather than Carnap's actual position. I argue that this is implausible, since deflationary readings of Carnap that immunise him against Beth's objection would make his position collapse into a form of Wittgenstein's radical conventionalism.

Part III looks beyond the position of *Logical Syntax* in different ways. In chapter seven I briefly discuss a number of contemporary and historically influential views in the philosophy of mathematics, including Frege's logicism, in order to see whether they face the problem Beth diagnosed as well. We will see that, in one way or other, these authors all need to rely on something like external mathematical facts or assumptions about what the one true logic is, and that hence Carnap cannot easily follow their example. Chapter eight revisits Quine, and contains a more detailed discussion of his positive views on truth in mathematics. Since Quine shares many of Carnap's convictions, such as empiricism and a broadly behaviouristic view of language, it is natural to expect that he will run into analogous problems. Against this I argue that Quine's attempt to make a mathematical ontology compatible with empiricism without relying on analyticity, while controversial in many ways, is not threatened by Beth's objection. Chapter nine finally leaves pure mathematics behind and investigates whether Beth's point teaches us anything about Carnap's account of ontology more broadly, including the ontology of the empirical world. I show that this is indeed so, since the core of the problem is Carnap's lack of interest in the role of natural language.

I conclude the thesis by briefly discussing how those sympathetic to Carnap's overall philosophical outlook should proceed. I distinguish between four broad strategies, but which one is the more fruitful remains to be seen.

Part I

Carnap and Mathematics

Chapter 1

Carnap's Internalism

It is convenient to have a label for Carnap's philosophy of mathematics, and I will therefore call his view *internalism*. In this chapter I introduce Carnap's internalism as it is presented in his *Logical Syntax of Language* and "Empiricism, Semantics, and Ontology". In the first half we will see that, from a contemporary perspective, it is tempting to classify Carnap as a mathematical Platonist, and then consider how the notion of a *linguistic framework* makes internalism distinctive. This theme is then continued in the second half, where I begin to explain what role the notion of *analyticity* plays in Carnap's philosophy. The chapter ends with a discussion of whether internalism is a form of linguistic conventionalism, a question we will also come back to in later chapters.

1.1 Platonism and Frameworks

1.1.1 Who's a Platonist?

In his paper "On What There Is" from 1948, Quine enumerates a number of "latter-day Platonists", and includes Carnap on this list (Quine 1948: 33).[1] This is somewhat surprising, for Platonism in mathematics is usually regarded as a position that comes with metaphysical commitments – whereas Carnap is known to as the arch-enemy of metaphysics. But things are not quite as simple. Carnap's most polemical anti-metaphysical writings are from the late 1920s and early 1930s (Carnap 1928, Carnap 1931), but if one looks into some of his works from the 1950s one can easily feel as if one is reading a paper by a contemporary metaphysician:

[1] In the reprint from 1953 Quine has removed this label from the paper, and Carnap is merely classified as a logicist (Quine 1953: 14).

> The term "concept" will be used here as a common designation for properties, relations, and similar entities For this term it is especially important to stress the fact that it is not to be understood in a mental sense, [...] but rather [as referring] to something objective that is found in nature and that is expressed in language by a designator of nonsentential form. (Carnap 1956b: 21)

> The fact that no references to mental conditions occur in existential statements [...] shows that propositions are not mental entities. Further, a statement of the existence of linguistic entities [...] must contain a reference to a language. The fact that no such reference occurs in the existential statements here, shows that propositions are not linguistic entities. The fact that in these statements no reference to a subject [...] shows that the propositions (and their properties, like necessity, etc.) are not subjective. (Carnap 1956a: 210f)

Did Carnap undergo a radical change of view in the intervening years? Some, including his close companion Otto Neurath, in fact thought so. Until the mid-1930s Carnap assumed that the rejection of metaphysical speculation also required one to reject semantic notions such as 'truth' and 'reference' altogether. In *Logical Syntax*, Carnap was still adamant that statements involving word-world relations – such as designation – are only admissible if understood in a loose sense, and should strictly speaking be replaced by meta-linguistic statements. The claim that the word 'daystar' designates the sun, for instance, is to be construed as follows:

> The word 'daystar' is synonymous with the word 'sun'. (Carnap 1937a: 289).[2]

Carnap changed his mind about the so-called doctrine of *pseudo-object sentences* after becoming acquainted with Tarski's theory of truth (Carnap 1963a: 60).[3] In "The Foundations of Logic and Mathematics" from 1939, he is happy to describe formal systems with straightforwardly semantic rules, such as

2 The use of synonymy – an intensional notion – may seem inappropriate in an account of the extensional notion of designation. In the context of this passage Carnap assumes that, at least for names, meaning and designation amount to the same thing.

3 In *Logical Syntax* Carnap objects to the notions of truth and falsehood on the grounds that they are "not proper syntactical properties" (Carnap 1937a: 216). It is actually quite difficult to understand why exactly Tarski's theory of truth convinced Carnap to change his position, for discussion see Ricketts 1996.

> 'mond' designates the moon (Carnap 1939: 9)

And since Carnap's languages also contain words for numbers, it was only natural that he also came to accept statements such as

> 'Five' designates a number. (Carnap 1956a: 217)

Neurath died in 1945, but he reacted strongly and negatively even to Carnap's initial acceptance of semantics. After reading Carnap's *Introduction to Semantics* from 1942, he wrote a letter with the following remark:

> Your Semantics copy did not arrive, I therefore tried to get one for a few days. I am just looking through the main chapters, particularly the chapters, you mentioned in your letter. I am really depressed to see here all the Aristotelian metaphysics in full glint and glamour, bewitching my dear fried Carnap through and through. (Neurath, Letter to Carnap from January 15, 1943 (Cat and Tuboly 2019: 570))

Neurath thus believed that by accepting semantics, Carnap had once again made the door for doing metaphysics wide open. The rest of Neurath's letter suggests that he thought that, once we allow word-world relations such as reference to count as meaningful, we end up being committed to a version of the correspondence theory of truth, according to which language is in the business of matching a structure reality has in itself. No doubt he would also have been worried by Carnap's assertion that 'five' designates a number, for that seems to commit us to a realm of abstract mathematical entities.

As we will see later, Neurath's interpretation of Carnap is largely mistaken. But his concerns are certainly understandable, for it is very tempting to describe Carnap's mature positions in terms that suggest certain metaphysical commitments. Herbert G. Bohnert, who was a student of Carnap, for instance characterises Carnap's philosophy of mathematics as follows:

> Gödel's platonism has often been said to be informally epitomized by his reference to "some well-determined reality, in which Cantor's (continuum) conjecture must be either true or false". But Carnap's platonism, was in a sense more extreme than Gödel. Abstract entities of every consistently imaginable (or unimaginable) sort existed, and every consistent set of axioms had many interpretations in the realm of abstract entities. (Bohnert 1975: 204f)

Carnap's attitude towards the abstract seems to have confused even some of his colleagues and students. In 1959, for instance, Carnap gave a talk about his treatment of theoretical terms in science at the Pacific APA.[4] In the discussion afterwards, Richard Montague and David Kaplan both asked questions about the background assumptions Carnap requires, in particular the need for higher-order logic and set theory with the axiom of infinity. They sound worried about these potentially controversial ontological commitments, but Carnap merely responds that ''you need a lot of assumptions, of course''.[5]

In a way Platonism is actually quite a good description of Carnap's position. For at least since Quine's work on ontology, it is common to identify Platonism with a willingness to quantify over numbers and sets, and Carnap was certainly happy to do that. Unlike Quine and Goodman he did not have much of an interest in trying to construct nominalist languages.[6] And he also always preferred to use classical logic and mathematics, having little sympathy for the philosophical motivations underlying constructivist approaches to mathematics. So it is quite appropriate to say that Carnap *talks* like a Platonist.

The crucial question then becomes whether talking like a Platonist is possible without any of the metaphysical commitments that have made nominalist or constructivist alternatives attractive to philosophers. As we will see in the next sections, the goal of Carnap's internalist program is to answer this question in the affirmative. Understanding how this is supposed to work exactly, and assessing whether his strategy *does* in fact work, will then occupy us for the rest of the thesis. Before we start it will be instructive to look at some senses in which Carnap *rejects* Platonism, however.

In his "Empiricism and Abstract Entities", which was a contribution to the volume on Carnap's philosophy edited by Paul Arthur Schilpp, Wilfrid Sellars raises an objection that is similar to Neurath's worries: Sellars thinks that once we allow reference-relations to obtain between numerals and numbers, we are committed to some sort of Platonic heaven (Sellars 1963). Carnap finds this objection unconvincing, for reasons that will become clearer later. What is especially helpful about his reply is that it reveals how Carnap himself conceived of

4 The talk can be found online (`https://www.youtube.com/watch?v=I7ebOwmCAqE`) and has been transcribed by Stathis Psillos (Psillos 2000b). This transcription does not include the brief discussion after the talk, however.

5 There is a transcript of the discussion in the Carnap Papers at the University of Pittsburgh (`https://digital.library.pitt.edu/islandora/object/pitt%3A31735061846840/viewer#page/50/mode/2up`).

6 Alspector-Kelly 2001 in fact argues that "Empiricism, Semantics, and Ontology" was a reaction to Goodman and Quine's "Steps Towards a Constructive Nominalism" (Goodman and Quine 1947).

Platonism:

> Relations of the causal type can indeed hold only among physical objects (or states or processes), not between a physical object and an abstract entity. It seems typical of Platonism, which both Sellars and I reject, that it speaks of relations of this causal type (called "commerce" or "intercourse" or the like) as holding between physical objects (or persons or minds) and abstract entities. (Carnap 1963b: 924f)

Carnap's assumption that Platonists *typically* believe in causal relations between persons and abstract objects seems rather questionable – indeed it is not so obvious whether *any* 20th century Platonist really endorsed such a bold hypothesis.[7] But it is useful to know that Carnap made this assumption, for it makes clear in what sense he took himself to reject Platonism: while he was happy to say that numerals refer to numbers, he denies that causal relations obtain between these entities.

The rejection of causal links to numbers is not the only metaphysical thesis about mathematics Carnap denies. As I already mentioned Carnap is happy to quantify over numbers, and thus counts as a Platonist by the standards of Quinean ontology. In principle Carnap has nothing against this classification either, but he is unhappy with the connotations the term 'ontology' brings:

> I should prefer not to use the word *'ontology'* for the recognition of entities by the admission of variables. This use seems to me to be at least misleading; it might be understood as implying that the decision to use certain kinds of variables must be based on ontological, metaphysical convictions. (Carnap 1956b: 43)

What is important to Carnap here is that there is no need to *justify* one's decision to quantify over abstract entities in any way. We will see why Carnap's holds this later on, but can already note that this attitude distinguishes him from other Platonists. For it is very natural to think that we should accept a theory that quantifies over numbers *because* there are numbers, and if we fail to do so we *miss out* on an aspect of reality. William Tait vividly describes a version of Platonism which fits the latter idea:

[7] Gödel's position is usually taken to be the 'most gung-ho version of platonism imaginable' (Potter 2001: 331), but as Potter shows it is doubtful whether Gödel actually wanted to explain mathematical knowledge by invoking a mysterious quasi-perceptual faculty of mathematical intuition.

> [...] mathematical practice takes place in an object language. But this practice needs to be explained. In other words, the object language has to be interpreted. The Platonist's way to interpret it is by Tarski's truth definition which interprets it as being about a model – a Model-in-the-Sky – which somehow exists independently of our mathematical practice and serves to adjudicate its correctness. (Tait 1986: 348)

On such a picture, the Model-in-the-Sky serves as a theory- and language-independent arbiter of mathematical truth, and could for instance be invoked in order to justify the acceptance of certain axioms over others as objectively correct. This is a way to think about mathematical objects that is diametrically opposed to Carnap's position, as we will soon be able to appreciate. For dialectical purposes it is useful to keep this alternative view in mind, however, and in the course of this thesis I will refer back to it from time to time. It is therefore convenient to label this position Model-in-the-Sky or *external* Platonism. With this background in mind let us now look at Carnap's internalist alternative.

1.1.2 Linguistic Frameworks

"Empiricism, Semantics, and Ontology" (ESO) starts with the following observation:

> Empiricists are in general rather suspicious with respect to any kind of abstract entities like properties, classes, relations, numbers, propositions, etc. They usually feel much more in sympathy with nominalists than with realists. [... On the other hand, the empiricist will probably] just speak about all these things like anybody else but with an uneasy conscience, like a man who in his everyday life does with qualms many things which are not in accord with the high moral principles he professes on Sundays. (Carnap 1956a: 205f)

The aim of ESO is thus to relieve empiricists of their guilty conscience and allow them to overcome their nominalistic scruples, by showing that using a Platonistic language "does not imply embracing a Platonic ontology but is perfectly compatible with empiricism and strictly scientific thinking" (Carnap 1956a: 206). If successful the account of ESO should thus be able to elucidate the question that arose in the previous section: namely why Carnap thinks that he can talk like a

Platonist without embracing metaphysics. And from a purely systematic standpoint this project is of great interest as well, since in contemporary philosophy quantifying over abstract objects is usually regarded as much more controversial than Carnap makes it out to be.

Carnap proceeds by relying on the notion of a *linguistic framework*, which he introduces as follows:

> If someone wishes to speak in his language about a new kind of entities, he has to introduce a system of new ways of speaking, subject to new rules; we shall call this procedure the construction of a linguistic *framework* for the new entities in question. And now we must distinguish two kinds of questions of existence: first, questions of the existence of certain entities of the new kind *within the framework*; we call them *internal questions*; and second, questions concerning the existence or reality *of the system of entities as a whole*, called *external questions*. (Carnap 1956a: 206)

There has been an extensive debate about the nature of Carnapian frameworks, but it is widely accepted that they are best understood as (fragments of) languages, which differ from natural language primarily in that they come with explicitly formulated *linguistic rules*.[8] What Carnap means by linguistic rules, however, is not straightforward. For expository purposes it is convenient to regard a framework as a formal theory with axioms and inference rules, even though we will see in the next chapter that Carnap actually has a broader understanding of what frameworks and their rules are.

Based on the notion of a linguistic framework Carnap then introduces the distinction between *internal* and *external* questions. By an *internal question* he means a question about whether some statement, such as "there are numbers" is true in a particular linguistic framework – i.e. given the axioms and inference rules the framework provides. *External questions* are more problematic, since here Carnap distinguishes two readings. In the quote above, Carnap says that they concern the reality of the system as a whole. What he has in mind is that external questions are not about whether some sentence is true *given* the framework, but instead concern whether the framework *itself* is true or false. One can read this as asking whether the axioms and inference rules provided by the

[8] For more on the nature of Carnapian frameworks see Broughton forthcoming. Eklund 2013 considers an alternative to the standard reading which construes Carnap as a *relativist*, and according to Flocke 2020 and Kraut 2020 Carnap should be understood as a *non-cognitivist* about ontology.

framework are correct, where correctness is understood in an *objective* sense. For the Model-in-the-Sky Platonist I introduced earlier, for instance, the relevant external question would be whether some framework we have constructed is such that the Model-in-the-Sky makes its axioms true and its rules truth-preserving – hence my suggestion to call this view *external Platonism*.

Call external questions of this kind *factual external questions*, since they presuppose that there are objective standards by which frameworks as a whole can be judged to be correct or incorrect, or at least better or worse. One crucial claim of ESO is that there is something *defective* about factual external questions, although Carnap is somewhat ambiguous about the nature of this defect. At some points he calls them "pseudo-statement without cognitive content" (Carnap 1956a: 214), which suggests that he regards them as meaningless. Another passage about an imagined dispute between a metaphysical Platonist and a nominalist sounds less damning:

> I cannot think of any possible evidence that would be regarded as relevant by both philosophers, and therefore, if actually found, would decide the controversy or at least make one of the opposite theses more probable than the other. [...] Therefore I feel compelled to regard the external question as a pseudo-question, until both parties to the controversy offer a common interpretation of the question as a cognitive question; this would involve an indication of possible evidence regarded as relevant by both sides. (Carnap 1956a: 219)

Here the problem seems to be not that external questions are literally meaningless, but rather that they are *unanswerable*, since we have no clue what counts as evidence for or against a particular answer.[9]

A second kind of external question are *pragmatic external questions*. Like factual external questions they concern frameworks as a whole, but instead of asking whether a framework is true or correct they concern whether it is *useful* for some practical purpose. Carnap thinks that there is nothing wrong with such questions, since there is obviously a sense in which it is better to adopt a framework which includes mathematics if we are interested in conducting calculations than to use one without. So there can be comparisons between frameworks, but, Carnap stresses, the choice of a framework can never be wrong but merely *in-*

[9] See Bradley 2018 for a reading of ESO that stresses the epistemological aspects of Carnap's critique of metaphysics. A contemporary view according to which metaphysical questions are meaningful but unanswerable can be found in Bennett 2009.

convenient.[10]

This is a good moment to address an issue that will likely occur to anyone with some background knowledge of logical empiricism and the Vienna Circle. Carnap is known to have endorsed a *verificationist* theory of meaning, and even though he became more liberal as time went on, he continued to endorse a form of empiricism that has few adherents nowadays. This raises the question whether Carnap's philosophy of mathematics *depends* on his verificationism and empiricism, or whether his position is relatively independent of these other commitments.

Some philosophers have had strong opinions concerning this matter. Carnap's rejection of factual external questions is not an innovation of ESO, but can already be found in *Logical Syntax*. Carnap had not yet adopted the terminology of frameworks and internal/external questions, but the thought that there is no objectively correct system of logic, and hence no need to provide epistemic reasons for the adoption of such a system, is expressed very forcefully:

> The fact that no attempts have been made to venture still further from the classical forms [of logic and mathematics] is perhaps due to the widely held opinion that any such deviation must be justified – that is, that the new language-form must be proved to be 'correct' and to constitute a faithful rendering of 'the true logic'. To eliminate this standpoint, together with the pseudo-problems and wearisome controversies which arise as a result of it, is one of the chief tasks of this book. (Carnap 1937a: xiv-xv)

> *Principle of Tolerance: It is not our business to set up prohibitions, but to arrive at conventions.* [...] *In logic there are no morals.* Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (Carnap 1937a: 51-52)

The so-called principle of tolerance has been widely discussed, and we will look at it in much more detail in chapter 4. What is worth considering now is that Hilary Putnam has claimed that Carnapian tolerance is not only is a *consequence* of the verification principle, but even *identical* to it:

[10] This of course raises the question in which (if any) framework the comparison between frameworks is supposed to proceed. For some illuminating discussions see Steinberger 2016 and Carus 2017.

> However, this principle of tolerance, as Carnap called it, *presupposes* the verification principle. For the doctrine that no rational reconstruction is uniquely *correct* or corresponds to the way things 'really are', the doctrine that all 'external questions' are without cognitive sense, *is* just the verification principle. (Putnam 1983: 191n)

I think that there is some truth in Putnam's claim. For Carnap needs to deny that we can make sense of the idea that a Model-in-the-Sky determines which framework is objectively correct, or at least he needs to say that we have no conception of what would count as evidence in favour or against such a hypothesis. And it seems that any argument to such a conclusion needs to make some broadly empiricist assumptions, such as that all evidence is empirical evidence we get through our senses, and that there is no special faculty of mathematical intuition.[11]

We should, however, not be tempted to overemphasise the role of Carnap's empiricist commitments for his philosophy of mathematics. It would be mistaken to think, as Putnam's remark may suggest, that anyone who rejects verificationism can with good conscience ignore Carnap's project as refuted. To see why it is useful to distinguish between the positive and the negative (or polemical) aspects of Carnap's position. The detailed reconstruction of the positive view will occupy us for long stretches of this thesis, but at this point it can be summed up as follows: according to Carnap we can make sense of mathematics *without* relying on a Model-in-the-Sky, since only framework-internal resources are needed in the end. And the success of this positive project is in principle compatible with the mere *existence* of something like a Model-in-the-Sky, as long as it doesn't play a theoretical role in the account of mathematics.

Now it is clear from Carnap's writings that he didn't merely want to be *agnostic* about whether there is a Model-in-the-Sky but deny that there is one, and the way he supports this *negative* thesis does plausibly rely on empiricism. But we should not spend too much time on this aspect of his view, for two reasons: first, the positive project is of great interest regardless of whether we also want to accept the negative thesis. Secondly, if the positive project actually succeeds, we can even argue for the negative thesis in a way that does *not* rely on empiricism, but on more innocuous considerations of theoretical economy. For if it really were true that we can give an account of mathematics that doesn't rely on something like a Model-in-the-Sky, why suppose that such a thing exists

[11] For more on the connection between Carnap's views on ontology and his empiricist epistemology, see Alspector-Kelly 2001 and Wilson 2011.

in the first place? This combination of views would be like accepting David Lewis's modal realism without having concrete possible worlds do any work in one's modal semantics, and could thus be set aside as a coherent but utterly unmotivated package.

Since the main aim of this thesis is to understand and evaluate Carnap's positive position, Carnap's verificationism will consequently not play a major role in the argumentation. Having settled this we can now continue the exposition of Carnap's internalism.

1.1.3 Internal Platonism

Even though Carnap himself does not use this terminology, it is apt to describe the view he espouses in ESO as *internal Platonism*: he thinks that we we can adopt a linguistic framework that is *Platonistic* since it allows us to quantify over numbers, but since the quantified statements are *internal* they are free of any problematic metaphysical commitments. The first step on the way to this internal Platonism is the adoption of a framework that has rules and axioms to talk about numbers:

> *The system of numbers*. As an example of a system which is of a logical rather than a factual nature let us take the system of natural numbers. The framework for this system is constructed by introducing into the language new expressions with suitable rules: (1) numerals like "five" and sentence forms like "there are five books on the table"; (2) the general term "number" for the new entities, and sentence forms like "five is a number"; (3) expressions for properties of numbers (e.g. "odd," "prime"), relations (e.g., "greater than") and functions (e.g. "plus"), and sentence forms like "two plus three is five"; (4) numerical variables ("*m*," "*n*," etc.) and quantifiers for universal sentences ("for every *n* . . . ") and existential sentences ("there is an *n* such that . . .") with the customary deductive rules. (Carnap 1956a: 208)

Adopting such a framework settles the internal question of whether there are numbers: since Carnap takes "five is a number" to follow from the rules of the framework, it is trivial that "there is a number" is true in the framework. The framework is hence Platonistic in the sense that it allows quantification over numbers.

We already saw some quotes in which Carnap went beyond saying that numbers and other abstract objects exist. In addition he for instance asserts that numerals *refer* to numbers, and that numbers are *mind-independent*. Although these sound like metaphysical claims at first sight, we can now see that for Carnap such statements are actually innocent since they are merely consequences of yet more linguistic rules.

I mentioned that by the time he wrote ESO, Carnap had accepted Tarski-style semantics as a successful formal explication of the notions of truth and reference. Consequently he is happy to use frameworks which include semantic notions such as 'truth' and 'refers', provided that when constructing the relevant framework we give rules for how to use these terms. If we set up a framework for semantics in the right way, we should for instance be able to derive

'*s*' designates *s*

for every singular term *s*. And if we have a combined framework which provides rules for mathematics *and* semantics, we will then be able to derive the following:

(1) Five is a number.

(2) 'Five' designates five.

Putting them together we then get the following claim, which appears to be a metaphysical thesis, but in the current context is just a harmless consequence of the rules for talking about numbers and reference:

(3) 'Five' designates a number.

So not only can Carnapian frameworks be Platonist in the sense of quantifying over numbers, they also allow us to say that numerals *designate* numbers.[12]

How about the claim that numbers, propositions, etc., are not subjective or mental, but rather mind-independent entities? For Carnap such claims are also consequences of the linguistic rules of a framework, albeit in a slightly more indirect way. The issue of mind-independence was already a matter of controversy among the logical empiricists in the 1930s, and it is instructive to consider the following passage in which Carnap discusses the claim that the existence and behaviour of stars is independent of the existence of minds and people. In this

[12] For a much more detailed exposition and defence of Carnap's treatment of ontology in ESO see Ebbs 2017a and Ebbs 2019. Ebbs addresses some issues I will not go into here, such as Carnap's surprising claim that some statements about the existence of physical objects are analytic.

example the predicate 'P$_3(t)$' means "there are no living beings in the world at time t", and 'P$_2$(t)' roughly means something like "the cosmos will be in a specific state P$_2$ at time t":

> (S$_5$) $P_3(t_0 + d) \supset P_2(t_0 + d)$
>
> S$_5$ may be taken as a convenient formulation of the following sentence discussed by Lewis and Schlick: "If all living minds (or: living beings) should disappear from the universe, the stars would still go on in their courses". [...] We have no well-confirmed predictions about the existence or non-existence of organisms at the time t_0 + d; *but the laws C of celestial mechanics are quite independent of this question*. Therefore, irrespective of its first part, S$_5$ is confirmed to the same degree as its second part, [...] and hence, as C. [...] Therefore I agree with the following conclusion of Schlick concerning the sentence mentioned above (though not with his reasoning): "We are as sure of it as of the best founded physical laws that science has discovered." (Carnap 1937b: 37f, my emphasis)

The salient point Carnap makes is as follows: we are justified to maintain that there is no connection between the existence of minds and the behaviour of planets, since the laws of celestial mechanics do not refer to human beings or their minds anywhere. In this sense stars and planets are thus mind-independent entities, unlike, presumably, social entities such as money or the state.

Understood in this way there is thus nothing objectionably metaphysical about assertions of mind-independence, and this is also what Carnap wants to stress when he describes properties and propositions as objective: the linguistic rules for proposition-talk merely specify that propositions are entities that are true or false, and, like the laws of celestial mechanics, do not invoke humans anywhere. Once again metaphysical-sounding claims, when rightly understood, boil down to mundane claims about linguistic rules.

Based on these clarifications we can now revisit the concerns Neurath and Sellars had about Carnap's acceptance of the reference relation. As I already suggested their line of reasoning seemed to be as follows: the reference relation is a word-world relation, hence if we accept that numerals refer to numbers we must assume that reality contains a realm of abstract objects to refer to, and so we immediately end up in metaphysical waters. Neurath was in general strongly opposed to the idea that the world consists of objects *in itself*, independently of our means of representation, in the style of Kant's *noumena*.[13] To Carnap's dismay Neurath even went so far as to complain about the existential quantifier for this reason:

> You remember, I always have been full of mistrust, as far as Russell's Existence symbol was concerned, and Russell [...] is just extending this start, which is closely related to your and Tarski's and Aristoteles' start: THERE EXIST SOMETHING IN ITSELF, this statement I thought is in a language not acknowledged by us (?) or by me (sure). (Neurath, Letter to Carnap from January 15, 1943 (Cat and Tuboly 2019: 571))

I think that Carnap can easily defuse this kind of worry though. He should agree that the reference relation is a world-word relation, but deny that this immediately leads to a commitment to anything like Kantian things in themselves. What Neurath (and also Sellars) fail to appreciate is that for Carnap semantics is just another linguistic framework like any other, and hence claims about words referring to objects are simply consequences of disquotation principles that are part of the linguistic rules governing 'refers' talk. A framework for semantics will typically include a rule that allows one to derive

'n' refers to n

for any singular term n.[14] Since numerals are singular terms, the conclusion that they refer to objects thus immediately follows. One might therefore reject semantics as unnecessary or impractical, but it cannot sensibly be dismissed as metaphysical, since, thanks to Tarski, there are clear and explicit rules for the use of semantic terms.[15]

For other philosophers the reference relation has a much more fundamental role: it is supposed to explain how language makes contact with reality in the

[13] See Stang 2016: section 6 for an overview.

[14] Carnap himself never considers the possibility of empty names, for such cases different disquotation principles would be needed.

[15] This internalist approach to truth and reference is central to Carnap's position, but it is difficult to formulate and hence easily misunderstood. Goldfarb and Ricketts for instance tend to describe Carnap as rejecting a "language-transcendent notion of empirical fact" and holding that "it is only given the apparatus of a linguistic framework that we can formulate the notion of the realm of empirical facts" (Goldfarb and Ricketts 1992: 65, see also Ricketts 1994). As Matti Eklund points out, it is tempting to misread this as amounting to no more than the trivial claim that we cannot talk about the world without using a language (Eklund 2012: 839n17).

first place.[16] In that case talking about metaphysics and the structure of the world in itself cannot easily be avoided. But this is not Carnap's view. For him, using a slogan from a slightly later debate, it is better to think of semantics as *just more theory*.[17] Even before he accepted the notions of truth and reference he makes a similar point concerning ostensions:

> Examples (2) are *ostensive* definitions; here the term is defined by the stipulation that the objects comprehended by the term must have a certain relation (for instance, congruence or likeness) to a certain indicated object; in linguistic formulation the ostension takes the form of a statement of the spatiotemporal position. *It is to be noted that, according to this, an ostensive definition likewise defines a symbol by means of other symbols (and not by means of extra-linguistic things).* (Carnap 1937a: 80, my emphasis)

Initially this seems confused, since ostensive definition seems to be *the* paradigmatic example of defining something by means of an extra-linguistic things, such as when one stipulates the referent of a new name by pointing at the intended referent. But I think what Carnap rejects here is a view of language that has sometimes been ascribed to Wittgenstein's *Tractatus*, according to which *mere signs* become meaningful symbols by associating them with objects through a mysterious act of ostension (Pears 1987: chapter 5).[18]

This last point raises an obvious and fundamental question: if for Carnap reference and ostension are not what brings language into contact with reality, how exactly do we manage to talk about the world then? I will come back to this issue in chapter 9, when we are in a position to consider the relationship between the linguistic frameworks discussed in this section and natural language. First,

[16] Bertrand Russell is plausibly a philosopher who falls into this category, as for him contact with the world is established through *direct acquaintance* with individual objects (Hylton 2006):

> The faculty of being acquainted with things other than itself is the main characteristic of a mind. Acquaintance with objects essentially consists in a relation between the mind and something other than the mind; it is this that constitutes the mind's power of knowing things. (Russell 1912: 22)

[17] Here I am alluding to Putnam's famous just-more-theory manœuvre in his critique of metaphysical realism. See Putnam 1980: 477 for one of Putnam's own formulations, and Taylor 1991 and Button 2013 for a comprehensive discussion.

[18] Whether this so-called *realist* interpretation of the Tractatus is what Wittgenstein actually had in mind is controverisal. For a helpful overview of the exegetical options see the first few sections of Bronzo 2017.

however, we need to consider another essential and controversial component of Carnap's philosophy of mathematics, namely the *analytic/synthetic distinction*.

1.2 Analyticity and Conventionalism

1.2.1 Truth and Linguistic Rules

In my exposition of Carnap's philosophy of mathematics I have so far focussed on the internal/external-distinction. It would be mistaken to interpret Carnap as arguing in the following way, however:

> Mathematical statements only make sense when read internally, therefore we must accept mathematics as true once we have adopted a mathematical framework.

This would be a fallacy since, according to Carnap, it is not only mathematical statements that need to be construed internally, but *all* statements whatsoever, including those about the empirical world. And if we adopt a framework in which we can talk about tigers, we obviously wouldn't expect or want that the truth values of all statements about tigers are settled by the mere adoption of the framework. Whether "there are no tigers in Africa" is true or false should depend on the way the world is after all, and determining its truth value requires empirical investigations.

There is thus an important difference between statements about the empirical world and statements about abstract objects. Whereas in the empirical case a linguistic framework merely determines the truth *conditions*, in the case of abstract object the framework directly determines the truth *values* of the relevant statements. And in order to draw this distinction, Carnap relies on the analytic/synthetic-distinction.[19]

Carnap expresses the general idea behind classifying some statements as analytic and others as synthetic as follows:

> In material interpretation, an analytic sentence is absolutely true whatever the empirical facts may be. Hence, it does not state anything about facts. [...] A synthetic sentence is sometimes true – namely,

[19] Could one accept the internal/external-distinction without also distinguishing between analytic and synthetic statements? This question has been discussed by Haack 1976 and George 2000, who suggest that Quine's position can be understood in this way. I come back to this issue in chapter 8, where we will see that Quine indeed rejects a certain external perspective on the relationship between language and the world.

> when certain facts exist – and sometimes false; hence it says something as to what facts exist. *Synthetic sentences* are the *genuine statements about reality*. (Carnap 1937a: 41 (§14))

That Carnap uses the notion of analyticity is probably one of the most well-known facts about his position, thanks to Quine's hugely influential "Two Dogmas of Empiricism" (Quine 1951). Even though it is common to regard Quine as the winner of this debate, evaluating the exact nature of their dispute is actually quite difficult, and I will postpone going into the details until chapter 3. For now it is important to keep in mind that Carnap's use of analyticity often differs from what contemporary philosophers have in mind, even though this is not always obvious on the surface. One target of Quine's objections, as well as Timothy Williamson's more recent arguments (Williamson 2007), for instance, is the claim that we can classify sentences of *natural languages* such as English or German into analytic and synthetic. But when Carnap invokes the analytic/synthetic distinction, he nearly always intends it to apply to linguistic frameworks with explicitly formalised rules. Furthermore, he explicitly states that whether we look at natural languages or linguistic frameworks makes an important difference for how to draw the distinction:

> Our explication, as mentioned above, will refer to semantical language-systems, not to natural languages. It shares this character with most of the explications of philosophically important concepts given in modern logic, e.g., Tarski's explication of truth. It seems to me that the problems of explicating concepts of this kind for natural languages are of an entirely different nature. (Carnap 1952: 66)

For now I will therefore only discuss the role of analyticity *within* Carnapian frameworks. In the quote above Carnap writes that analytic sentences are "absolutely true whatever the empirical facts may be", so let us start by considering in what sense they are immune from revision in light of incoming evidence. This is important since the idea that adopting a mathematical framework settles all there is to say about mathematics only makes sense if, once a such a framework is adopted, there are no empirical considerations that could speak in favour or against accepting certain mathematical statements. In other words, we need some explanation of why, within a framework, there can be empirical evidence for a sentence like "there are tigers in Africa" but not "there are infinitely many prime numbers".

In order to appreciate how this works, we need to understand how Carnap deals with evidence in general. There are two crucial tenets. Firstly, for Carnap all evidence is *empirical* evidence, and secondly, evidence is *linguistically mediated*. The first point should be clear enough, but the second requires further elaboration. Frameworks which can talk about the empirical world require there to be a class of *protocol-sentences*:

> Syntactical rules will have to be stated concerning the forms which the *protocol-sentences*, by means of which the results of observation are expressed, may take. (Carnap 1937a: 317)

These protocol-sentences are then connected to events in the empirical world by treating some of the predicates that occur in them as observable, which amounts to the following:

> A predicate 'P' of a language L is called *observable* for an organism (e.g. a person) N, if, for suitable arguments, e.g. 'b', N is able under suitable circumstances to come to a decision with the help of few observations about a full sentence, say 'P(b)', i.e. to a confirmation of either 'P(b)' or '~P(b)' of such a high degree that he will either accept or reject 'P(b)'. (Carnap 1936: 454f)

With respect to their relation to protocol-sentences, analytic sentences then have the following property:

> If, however, we assume that every new protocol-sentence which appears within a language is synthetic, there is this difference between an L-valid, and therefore analytic, sentence $\mathfrak{S}_1$ and a P-valid sentence $\mathfrak{S}_2$, namely, that such a new protocol-sentence – independently of whether it is acknowledged as valid or not – can be, at most, incompatible with $\mathfrak{S}_2$ but never with $\mathfrak{S}_1$. (Carnap 1937a: 318f (§82))

The P-rules Carnap mentions are supposed to be laws of nature, but we can set these aside here in order not to complicate the discussion. The crucial point is that Carnap identifies analytic sentences with logically valid (L-valid) sentences, and claims that they cannot stand in conflict with protocol-sentences, which in turn entails that there can be no empirical evidence that speaks *against* an analytic sentence. Why exactly is this the case? Carnap first assumes that protocol-sentences are synthetic, which just means that they are not analytic, i.e. L-valid. For an L-valid sentence to be incompatible with a synthetic sentence *S*, it would have to entail $\neg S$. Since the negation of a synthetic sentence is also synthetic, there would thus need to be an *analytic* sentence that entails a *synthetic* sentence. But that is impossible, since all the consequences of an L-valid sentence must be L-valid, and hence analytic, themselves.

The effect of this is that analytic sentences become immune to revision, as Carnap himself describes it here:

> Since the truth of an analytic sentence depends on the meaning, and is determined by the language rules and not the observed facts, then an analytic sentence is indeed "unrevisable" in another sense: it remains true and analytic as long as the language rules are not changed. The attribution of truth values to synthetic sentences changes continually, induced by new observations, even during a period in which the logical structure of language remains unchanged. A revision of this sort is not possible for the analytic sentences. (Carnap 1990: 432)

And this explains why for analytic sentences, there is no gap between accepting a framework that contains them, and accepting them as true: the only factor that might prevent us from regarding them as true would be empirical evidence to the contrary, but by design there can be no such thing.

We have now seen what kind of work the distinction between analytic and synthetic sentences is supposed to do within a Carnapian framework. What remains to be seen, however, is how the extension of the predicate 'analytic' is determined – i.e. how exactly the rules of the framework settle *which* sentences are analytic. In ESO, Carnap writes that in order to answer questions about mathematics, we need to employ "logical analysis based on the rules for the new expressions" (Carnap 1956a: 209), which sounds straightforward. But we will soon see that Carnap's aim is to declare *all* purely mathematical sentences to be either analytic or contradictory, and this is not a trivial task at all given the limitative results Gödel proved in the 1930s. Carnap's proposal for dealing with Gödel's incompleteness theorems will therefore be the focus of the next chapter, but before we come to that I want to situate Carnap's position in the established theoretical landscape some more. We started by looking at Carnap's relation to Platonism, and now is a good time to ask whether his philosophy of mathematics should be classified as a form of *linguistic conventionalism*.

1.2.2 Is this Conventionalism?

Linguistic conventionalism about an area of discourse is the thesis that, in some sense, the statements of that area do not describe reality or state facts, but are rather true in virtue of, or can be explained by, certain linguistic conventions. Linguistic conventionalism has been historically popular when applied to logic and mathematics, especially among advocates of empiricism. This is not surprising, since on the surface logical and mathematical knowledge seems quite different from empirical knowledge. While the latter is based on sense experience and concerns contingent matters of facts, the former is often regarded as a priori and concerns necessary truths and falsehoods. Some empiricist, such as John Stuart Mill and later Quine, denied that there is a sharp dividing line between logic and mathematics on the one hand, and empirical matters on the other. Among the logical empiricists, however, the favoured strategy was to retain the intuition that there is something distinctive about logic and mathematics, but to try to explain this difference without compromising empiricism.

A. J. Ayer sums up the idea behind linguistic conventionalism succinctly in his influential *Language, Truth, and Logic*. With regards to the "the a priori propositions of logic and pure mathematics" he writes that

> [...] I allow [them] to be necessary and certain only because they are analytic. That is, I maintain that the reason why these propositions cannot be confuted in experience is that they do not make any assertion about the empirical world, but simply record our determination to use symbols in a certain fashion. (Ayer 1936: 31)

The underlying thought is that by classifying logic and mathematics as analytic, and hence as obtaining in virtue of linguistic conventions, their necessity and a priority becomes palatable even for an empiricist. For to call the laws of logic necessary means nothing more than that they follow from certain rules for using signs we have adopted as a convention, not that there are certain features of the world that make these truths necessary. And they are a priori since it is assumed that we can know what linguistic rules we are following without using observation. It is hence common to find formulations similar to Ayer's in the 1930s, not only in the writings of the Vienna Circle but also expressed by American philosophers such as C. I. Lewis.[20]

Carnap seems to fall squarely into the camp of linguistic conventionalists. We already saw that he wants to classify all of logic and mathematics as analytic,

[20] See Lewy 1940 for a collection of representative quotes.

and regards analytic sentences as true in virtue of framework rules. In a paper from the mid-1930s, he also stresses that he considers the formal sciences of logic and mathematics to be empty of content:

> In adjoining the formal sciences to the factual sciences *no new area of subject matter* is introduced, despite the contrary opinion of some philosophers who believe that the "real" objects of the factual sciences must be contrasted with the "formal", "*geistig*" or "ideal" objects of the formal sciences. *The formal sciences do not have any objects at all*; they are systems of auxiliary statements without objects and without content. (Carnap 1953: 128)

If Carnap is a linguistic conventionalist, then this raises some troubling questions. For nowadays conventionalism about logic and mathematics is not a popular view anymore, since it is usually thought that there are compelling and near-insurmountable objections to the position. In the following I will briefly introduce the two most well-known arguments against conventionalism: Quine's circularity argument from his "Truth by Convention", and what I call the Master Argument against the very idea of conventionalism. Then I will discuss whether these arguments in fact apply to Carnap's own position.

The gist of Quine's argument can be summed up as follows: if logic is true by convention, then each logical truth must be stipulated to be true. There are infinitely many logical truths, however, whereas no one can make infinitely many stipulations. The only way out, so Quine, would be to stipulate infinitely many sentences to be true by using some finite mode of expression:

> It would appear that we sit down to a list of expressions and check off as arbitrarily true all those which, under ordinary usage, are true statements involving only our logical primitives essentially; but this picture wanes when we reflect that the number of such statements is infinite. If the convention whereby those statements are singled out as true is to be formulated in finite terms, we must avail ourselves of conditions finite in length which determine infinite classes of expressions. (Quine 1949: 262f)

And at first sight that doesn't seem to be a grave difficulty, for we could apparently use the following stipulations:

(I) Let every instance of the following schema be true: $\ulcorner \phi \to \phi \urcorner$.

(II) If a statement ϕ and a statement $\ulcorner\phi \rightarrow \psi\urcorner$ are true, then let ψ be true as well.

Quine thinks that this move would defeat the whole purpose of logical conventionalism, however. For it seems that in order to establish that, for instance,

$$(p \rightarrow p) \rightarrow (p \rightarrow p)$$

is a logical truth, one already needs to *use* some logic. It needs to be shown that

"$(p \rightarrow p) \rightarrow (p \rightarrow p)$" is an instance of (I),

and then we can infer that

"$(p \rightarrow p) \rightarrow (p \rightarrow p)$" is true.

But the first step requires the use of something like universal instantiation, and the second step is plausibly just reasoning according to modus ponens. Quine sums up the issue as follows:

> In a word, the difficulty is that if logic is to proceed mediately from conventions, logic is needed for inferring logic from the conventions. (Quine 1949: 271)

His conclusion is thus that logical conventionalism fails to provide a *non-circular* explanation of the nature of logic, for there appears to be no way to account for the logical truth of infinitely many sentences without already presupposing at least a limited number of logical principles.

What is notable about Quine's argument is that he grants that we can stipulate *individual* sentences to be true, and the problem only arises due to the infinity of logical truths. The second major argument against conventionalism, however, goes further and targets the fundamental assumption that truth in virtue of conventions is a coherent notion at all – hence I call it the Master Argument. Its exact origins are not so clear, but Stephen Yablo has attributed it to a book by Casimir Lewy from 1976:

> Near the middle of his book *Meaning and Modality*, Casimir Lewy takes up the theory that "necessary propositions. . . ' owe their truth to' linguistic conventions." All that conventions can do, he protests, is help to determine what a sentence *says*, or what proposition it expresses; whether the proposition holds true is then another question, to which rules of usage are quite irrelevant. [...] With doubtful historical accuracy I will call this the Lewy point. (Yablo 1992: 878)

Nowadays this kind of argument is probably more strongly associated with Paul Boghossian's paper "Analyticity Reconsidered" from 1996, which stresses the same general point: everyone can agree that conventions determine that a sentence *s* means that *p* – but whether *p* is *true* is then a separate question, and in most cases conventions are not a plausible candidate for explaining *why p* is true:

> Are we to suppose that, prior to our stipulating a meaning for the sentence
>
> Either snow is white or it isn't
>
> it wasn't the case that either snow was white or it wasn't? Isn't it overwhelmingly obvious that this claim was true before such an act of meaning, and that it would have been true even if no one had thought about it, or chosen it to be expressed by one of our sentences? (Boghossian 2017: 583)

For the thesis that logic and mathematics are true in virtue of conventions to be philosophically interesting, it should neither amount to the uncontroversial claim that conventions fix the meaning of logical vocabulary, nor to the crazy claim that we literally made it the case that logical laws obtain. The very idea of conventionalism is thus in jeopardy, since, so Boghossian, it is hard to see what alternative to these two unpalatable options remains.

1.2.3 Carnap's Deflationism

Quine's circularity problem and the Master Argument have been very influential, and as a consequence there are hardly any card-carrying conventionalists around today.[21] The important question for us is whether these arguments also discredit Carnap's internalism. I think the answer is *no*, for a reason that many Carnap scholars have stressed in recent decades: namely that Carnap's overall approach to philosophy is often *deflationary*. He doesn't see himself as *answering* traditional philosophical questions, but rather sets them aside as too unclear, and proposes to replace them with more tractable kinds of enquiries. In this section we will see how this general point applies to the issue of conventionalism.

[21] Things might change, however, as a book-length defence of conventionalism has just appeared (Warren 2020b). I briefly discuss the relation of Warren's project to Carnap's position in section 7.4.

I characterised linguistic conventionalism as the thesis that conventions can *explain* logical and mathematical truths, or that logical and mathematical statements are true *in virtue of* linguistic conventions. These formulations are common in the literature, but without further elaboration the notions of *explanation* and *true in virtue of* are hardly unproblematic. A number of commentators have consequently argued that Carnap would reject these notions outright, and should for this reason not be classified as a conventionalist:

> We can now appreciate the deflationary character of Carnap's philosophy of mathematics. Gödel's conventionalist target contrasts empirical truth with the truth conferred by conventional stipulation. Carnap rejects this contrast; he rejects any thick notion of truth-in-virtue-of. (Ricketts 2007: 211)

> Contrary to what many believe, [Carnap] rejects the confused thesis that our acceptance of logical and mathematical sentences somehow guarantees that those sentences are true. (Ebbs 2017b: 25)

This may seem surprising, since clearly Carnap does think that there is *some* sense in which the truth values of analytic sentences are settled by the linguistic rules of frameworks. The deflationary interpreters don't want to deny this either, for Goldfarb and Ricketts are perfectly happy to describe Carnap's position as one according to which "mathematical truths [...] flow from the adoption of the metalanguage" (Goldfarb and Ricketts 1992: 71).[22] The point they want to make is rather that we need to be careful not to read the claims Carnap's makes in a too metaphysically loaded way. And it is indeed important to keep this in mind, for in light of the resurgence of interest in notions such as grounding and metaphysical explanation it is tempting to read Carnap as trying to answer questions he himself would probably have rejected as being ill-formed altogether.

The deflationary character of Carnap's position enables him to set aside Quine's circularity objection and the Master Argument. In the case of Quine this is relatively straightforward. His complaint was that conventionalism cannot give a *non-circular* account of logical and mathematical truth, but it is clear that giving an explanation of this kind cannot have been Carnap's aim in the first place. In the next chapter we will consider Carnap's account of mathematical truth in more detail, and see that he *uses* mathematics in the relevant definitions. It can hardly have escaped his notice that there is something circular about this

[22] See also Ebbs 2017b: 26 for an analysis of the notion of "determination".

approach, and so there must be some truth to Goldfarb and Ricketts' interpretation according to which Carnap had no interest in giving a non-circular account of logic and mathematics:

> Carnap's position contains a circle, or, better, a regress: mathematics is obtained from rules of syntax in a sense that can be made out only if mathematics is taken for granted (in the metalanguage). Therefore, no full exhibition of the syntactical nature of mathematics is possible. This is not lethal, however, insofar as the structure of Carnap's leaves no place for the traditional foundational questions that such an answer would certainly beg. (Goldfarb 1995: 330, see also Goldfarb and Ricketts 1992: 71)[23]

How about the Master Argument though, which seems to show that there is no sense in which a sentence can be true by convention that is both interesting and plausible? It is intriguing to see that on one occasion Carnap himself argues in an analogous way to Lewy and Boghossian:

> [...] the logical truth of the sentence "all black dogs are dogs" is not a matter of convention [...]. Once the meanings of the individual words in a sentence of this form are given (which may be regarded as a matter of convention), then it is no longer a matter of convention or of arbitrary choice whether or not to regard the sentence as true; the truth of such a sentence is determined by the logical relations holding between the given meanings. (Carnap 1963b: 916)

The thought here is that if one considers the sentence "all black dogs are dogs" together with its actual meaning – which one could also describe as the proposition expressed by it – then its truth is no longer a matter of convention. And this suggests that Carnap himself holds the moderate view that all that conventions do is to fix the meanings of sentences, while truth is another matter.

I think that this description is partly correct, but partly misleading as well. Boghossian seems to assume that moderate conventionalism is philosophically uninteresting since it fails to draw a distinction between empirical truth, on the one hand, and logical and mathematical truth, on the other. In both cases sentences need to be connected to propositions by conventions, after all, and hence it is hard to see why truth in the one case would be more a matter of convention

23 For a discussion of whether Quine *himself* had Carnap as his target in mind when writing "Truth by Convention" see Ebbs 2011.

than the other. And while this sounds very compelling, I think that the underlying conception of sentences and their relation to propositions is not as innocent as it seems. Jared Warren expresses this well in the following passage:

> This picture seems to assume that there are the propositions, somewhere out there all arrayed. They toil not, they spin not; they are timeless and forever. We corporeal beings work not with propositions but with sentences. Our conventions generate meaningful sentences simply by attaching them to particular propositions—like price tags at the grocery store. (Warren 2015b: 90f)

This way of thinking about propositions is very different from Carnap's. While he is happy to grant that there are such entities, he conceives of them in an internalist manner, namely as being constituted by the rules of linguistic frameworks:

> *The system of propositions*. New variables, "*p*," "*q*," etc., are introduced with a role to the effect that any (declarative) sentence may be substituted for a variable of this kind; [...]. Further, the general term "proposition" is introduced. "*p* is a proposition" may be defined by "*p* or not *p*" (or by any other sentence form yielding only analytic sentences) . Therefore every sentence of the form ". . . is a proposition" (where any sentence may stand in the place of the dots) is analytic. (Carnap 1956a: 209f)

This is important to note since it shows that, for Carnap, there is a sense in which sentences are *prior* to propositions. For while every Carnapian framework will consist of sentences, propositions are merely useful but optional additions, which, furthermore, are characterised by reference to sentences. And this deflationary conception of propositions makes room for a reply to the Master Argument that was not so obvious before we explicitly discussed their status.

I think that, when pressed, Carnap should say the following: there is clear a sense in which it is *false* that linguistic conventions make it the case that propositions obtain. For, in analogy with the earlier discussion of the mind-independence of stars, within the frameworks Carnap proposes it will be false to claim that, for instance, 2+2=4 holds because of a convention. But this is not to say that all propositions – both logico-mathematical and empirical – are alike in all relevant respects, contra to what Boghossian suggests. For linguistic frameworks can be set up in such a way that logical and mathematical sentences are analytic, which guarantees their truth. And since propositions are defined in terms of sentences, there is a derived sense in which those propositions can also said to be true in virtue of linguistic conventions. It is, however, preferable to avoid this way of speaking, since it misleadingly suggests that we literally make the laws of logic obtain. A more innocuous formulation, then, would be to say that we can set up linguistic frameworks in such a way that some sentences – namely the analytic ones – are *guaranteed to express true propositions*.[24] Naturally this claim can only be expressed from the perspective of someone who stands outside of the framework whose properties are being discussed, and hence there is no tension with the earlier denial of conventionalism from within the framework.

In this section I have argued that, thanks to his deflationary conception of philosophy, Carnap can answer the two most influential arguments against linguistic conventionalism. In a way the strategy pursued here raises more questions than it answers, however. For it is easy to worry that through rejecting more substantial explanatory projects, Carnap's own position has become so deflationary that nothing of interest remains.[25] Less drastically speaking, there is at least a serious questions as to what sort of positive claims Carnap *did* want to make, and why he thought these to be important. The next chapter will start to answer some of these concerns by looking at Carnap's conception of linguistic rules more closely, with a focus on his treatment of Gödel's incompleteness results.

[24] This idea also plays a role in Gillian Russell's recent attempts to rehabilitate truth in virtue of meaning (Russell 2008).

[25] Quine indeed ends his "Truth by Convention" by raising the worry that conventionalism becomes an "idle label" without any explanatory force (Quine 1949: 273). I will come back to similar considerations in Quine's later writings in section 3.2.

Chapter 2

Gödel and Incompleteness

In the previous chapter we have seen that the aim of Carnap's internalism is to allow the use of mathematics while avoiding any commitment to metaphysical forms of Platonism. This is to be achieved by construing all true mathematical statements as *analytic*. But since analyticity is spelled out in term of what follows from linguistic rules, one might think that Carnap's project is doomed from the start. By Gödel's incompleteness theorems there is no formal system such that for every mathematical sentence either it or its negation is derivable. What is the status of these undecidable sentences then? Carnap seems forced to either hold that they have no truth value at all, or that they are *synthetic* sentences like those about the empirical world.

Carnap thought that he had a way out of this apparent dilemma, and in this chapter I will describe his strategy of overcoming Gödel's limitative results. The main idea is to broaden one's conception of *linguistic rules* in such a way that non-recursive and infinitary rules can be used to characterise what the analytic sentences are. Even though this strategy was a direct result of suggestions by Gödel, Gödel himself was not at all convinced by it, and argued that the second incompleteness theorem undermines Carnap's approach after all. In the second half of the chapter I will review Gödel's *argument from consistency*, and discuss some replies to it that have been given on Carnap's behalf.

2.1 Linguistic Rules Revisited

2.1.1 Syntactic Rules and Their Limits

Let us begin by stating Carnap's goal in a more precise way. Call a sentence *analytic* if it follows from the rules of a linguistic framework, *contradictory* if its negation follows, and *synthetic* if it is independent of the rules. What Carnap wants to account for is the following claim (Carnap 1937a: 116):

> Classical Mathematics
> Every purely mathematical sentence, i.e. every sentence which only contains mathematical vocabulary, is either analytic or contradictory.

This is supposed to capture the classical idea that every purely mathematical sentence is either determinately true or determinately false. Note that Carnap does not want to deny that other, more constructivist or formalist conceptions of mathematics, according to which bivalence does not hold, are *possible*. He would have no objection against someone who adopts a framework for doing intuitionist mathematics, for instance. But what Carnap denies is that there is something *philosophically problematic* about classical mathematics, and that we therefore *have to* opt for a constructivist alternative. And in order to achieve this, he needs to establish that *there is* a linguistic framework which captures classical mathematics in the way I described, even though this is not *the only* legitimate framework.[1]

Having clarified Carnap's aim, we now need to take a closer look at what it means for sentences to follow from the linguistic rules of a framework. I already mentioned that, based on ESO, it is very tempting to identify a framework with the kind of formal theories logicians typically study, i.e. a set of axioms and inference rules. Putting these ideas together, it is then natural to spell out *analyticity* in the following way:

> A sentence S is analytic iff it can be derived from the axioms that are constitutive of the framework using the inference rules the framework provides.

This conception is intuitive, but immediately leads to trouble. If frameworks are formal theories in this sense, then it seems that the framework of arithmetic just is something like Peano arithmetic (PA). So suppose we identify analyticity with what is derivable from the axioms of PA:

> Syntactic Proposal
> ϕ is analytic **iff** PA $\vdash \phi$.

[1] We will later see that the dialectical situation is actually a bit more complicated, but in order to keep the exposition of Carnap's position straightforward I postpone further discussion to chapter 6.

At this point Gödelian incompleteness strikes. For if we assume that PA is in fact consistent, one can show that there is some sentence G, formulated in arithmetic vocabulary, such that PA neither proves G nor $\neg G$. From the way we characterised analyticity it follows that, on the Syntactic Proposal, G counts as *synthetic*. And this is just the kind of result we didn't want in a framework for classical mathematics.[2]

What makes this case worrisome is that PA is just a special case. In its most general form, Gödel's incompleteness theorems show that there is no theory T which has all of the following properties:

(1) T is consistent

(2) T is recursively formalised

(3) T is strong enough to do basic arithmetic

(4) T is complete, i.e. for any T-sentence ϕ, either T proves ϕ or T proves $\neg\phi$

And this seems to preclude the possibility of identifying analyticity with what it provable in *any* theory, not just PA, as long as we stick to the requirement that every mathematical sentence is either analytic or contradictory.

The Syntactic Proposal of analyticity is useful to illustrate this challenge, but, ever since the 1930s, Carnap's actual conception of analyticity was more sophisticated. For he was very aware of Gödel's incompleteness results; Carnap indeed seems to have been one of the first people Gödel talked to about this matter in a Vienna coffee shop, and they frequently discussed the philosophical implications afterwards.[3] Carnap thought that he had found a way to overcome these limitations, and we will turn to his own proposal now.

In order to understand Carnap's strategy, it is useful to consider condition (2) from above – that theories are recursively formalised – in more detail. Panu Raatikainen helpfully sums up what kinds of formal theories the incompleteness theorems apply to as follows:

> Roughly, a *formal system* is a system of axioms equipped with rules of inference, which allow one to generate new theorems. The set of

[2] For a philosophically informed discussion of Gödel's theorems and other limitative results see Smith 2013.

[3] In a diary entry from 26th August 1930, Carnap writes: "Gödel's discovery: incompleteness of the system of *PM*; the difficulty of a consistency proof" (Carnap forthcoming: 818, my translation). For more background on the early interactions between Carnap and Gödel see Köhler 2002, Goldfarb 2005, and Awodey and Carus 2010.

> axioms is required to be finite or at least decidable, i.e., there must be an algorithm (an effective method) which enables one to mechanically decide whether a given statement is an axiom or not. If this condition is satisfied, the theory is called "recursively axiomatizable", or, simply, "axiomatizable". The rules of inference (of a formal system) are also effective operations, such that it can always be mechanically decided whether one has a legitimate application of a rule of inference at hand. Consequently, it is also possible to decide for any given finite sequence of formulas, whether it constitutes a genuine derivation, or a proof, in the system – given the axioms and the rules of inference of the system. (Raatikainen 2020: section 1.1)

This is important to keep in mind, since one can also talk about formal systems or theories in a looser sense. Consider True Arithmetic (TA): the theory whose axioms are all the true arithmetical sentences. TA is not incomplete, since – assuming arithmetical truth is classical – for every arithmetical sentence S, either S is true and is hence an axiom of TA, or S is false, and hence the negation of S is an axiom. TA is not a formal system in the sense of Raatikainen, however, i.e. it is not recursively formalised, since there no effective way to list its axioms. For theories like TA, there is no way to represent the notion of 'provable-in-the-theory' within the language of arithmetic, which is what any proof of Gödel's theorems relies on.

At a high level of abstraction, one can understand Carnap's approach as characterising analyticity by means of a non-recursive theory. To make this more concrete, let us look at how Carnap sets things up in *The Logical Syntax of Language* (Carnap 1937a). In the book he introduces two formal theories called language I and II, of which the former is a version of primitive recursive arithmetic, whereas the latter is a version of the simple theory of types with higher-order quantifiers.[4] In the following I will continue to talk about PA, however, since this theory is very familiar and hence a useful example. No harm is done by this, since the typed nature of language II is not important when it comes to Carnap's strategy of resolving incompleteness.

By Gödel's second incompleteness theorem, one example of a sentence independent from the axioms of PA is Con_{PA} – the sentence which encodes the claim that PA is consistent. It is possible to define a predicate Pr_{PA} that represents PA's provability-relation within the language of arithmetic, such that $Pr_{PA}(\ulcorner\Sigma\urcorner, \ulcorner\phi\urcorner)$

[4] At that time Carnap had not adopted the terminology of ESO yet, but what he calls 'languages' in *Logical Syntax* are in effect linguistic frameworks.

is true if and only if Σ is a PA-proof of ϕ.[5] Letting '⊥' stand for some arbitrary contradiction, the sentence Con_{PA} is then an abbreviation of the following all-quantified claim:

$$Con_{PA} =_{def} \forall x \neg Pr_{PA}(x, \ulcorner \bot \urcorner)$$

In *Logical Syntax*, Carnap likewise picks out Con_{PA}[6] as an example of a sentence that is *analytic* despite not being *derivable*. In order to see why he is allowed to say this, Carnap's definition of analyticity for quantified statements of the form '$\forall x P_1(x)$' is relevant. It is given after analyticity has already been defined for monadic predications, and relies on an infinite class of accented expressions or numerals 0, $0'$, $0''$, ... In order to determine whether '$\forall x P_1(x)$' is analytic, one needs to

> refer for instance from '$P_1(x)$' to the sentences of the infinite sentential class $\{$'$P_1(0)$','$P_1(0')$','$P_1(0'')$',...$\}$. In this manner, the numerical variable is eliminated. (Carnap 1937a: 106 (§34c))

This is then supposed to settle the analyticity of Con_{PA}. For all instances of this universal generalisation – i.e. $\neg Pr_{PA}(0, \ulcorner \bot \urcorner)$, $\neg Pr_{PA}(0', \ulcorner \bot \urcorner)$, ... – *are* provable in PA, and thus analytic. The definition of analyticity in the metalanguage thus has the effect that if all instances of some formula $\phi(x)$ are true in virtue of the framework rules, then so is the universal generalisation $\forall x \phi(x)$, which in turn settles the truth value of Con_{PA}.

The general idea behind this move is that while PA+Con_{PA} and PA+$\neg Con_{PA}$ are both consistent, the theory that results from adding $\neg Con_{PA}$ only has *non-standard models* – i.e. models which are not isomorphic to the natural numbers, but contain additional non-standard elements. Con_{PA}, on the other hand, is true in the standard model. The truth of Con_{PA} can thus be secured if we are in a position to set non-standard models aside as irrelevant. And indeed, Carnap's definition from above only works if the infinite stock of numerals '0', '0′', '0″', ... he refers to is isomorphic to the natural numbers, and thus comprises only standard numerals. He clearly assumes this to be so, but the discussion of Beth's arguments in chapter 4 will show that this presupposition is not without its problems.

Earlier I already mentioned Carnap's transition from his syntactic to his semantic period. *Logical Syntax* is still part of the syntactic period during which

[5] The Quine-corners '⌜ ⌝' map an expression in the language of arithmetic to its Gödel-number.
[6] Or rather the consistency sentence of language II, but I will not flag this every time.

he rejected semantic notions such as truth and reference. The acceptance of semantics does not make a substantial difference for the issues under discussion, however. The definition of analyticity we just saw, even though *called* syntactic, is in effect a notational variant of a Tarskian truth-definition (Coffa 1987, Koellner ms). Carnap invokes a particular set of numerals at a crucial juncture of his definition, and this quasi-semantic approach just became more explicit once he officially accepted semantics. In a later summary of his views, he for instance suggests that in the metalanguage we should just describe the intended interpretation for the object language under consideration (Carnap 1963b: 900f).

The sentence Con_{PA} is just one particular example here, even though an especially interesting one. But the approach generalises, for every arithmetical sentence is either true or false in the standard model of arithmetic $\mathbb{N}$. If one wanted to sum up Carnap's conception of analyticity in *Logical Syntax* in one sentence, the following proposal would be the most accurate:

> Semantic Proposal
> ϕ is analytic **iff** $\mathbb{N} \vDash \phi$.

The Syntactic Proposal only relied on the syntactic notion of derivability (symbolised as '$\vdash$'), and hence suggested that Carnapian frameworks can be identified with a formal language plus a deductive system. The Semantic Proposal shows this interpretation to be misguided. It uses the semantic notion of *truth in a model*, symbolised as '$\vDash$', and understood in this way a linguistic framework must also include a particular semantics.

In one sense this strategy is not that surprising: it is pretty much the norm to interpret Gödel's incompleteness results as showing that there are *true* but *unprovable* sentences in mathematics – where truth is in effect understood as truth in the standard model. So far Carnap's approach is in line with the mainstream. What is surprising, however, that he thinks that the moves he makes are compatible with conceiving of mathematical truth as a matter of linguistic rules, for it may seem that relying on a particular model is rather to *give up* on this idea. In the next section we will therefore need to investigate this issue in more detail.

2.1.2 The Nature of Frameworks

The most important result of the previous discussion is that Carnap's conception of what linguistic rules are is very generous. It has been noted before that some of Carnap's moves are bound to surprise contemporary readers. He frequently suggests that he thinks of analyticity as a kind of logical truth, for instance, which does not easily fit the proposal we just encountered. Logical truth is usually spelled out as truth in *all* models, after all, rather than truth in *one* particular interpretation. Frost-Arnold thus rightly stresses that Carnap has an unusually broad conception of what can rightfully count as analytic:

> This difficulty is solved by recognizing that Carnap does not characterize analytic truth as truth-in-all-interpretations. Analytic truth for Carnap, as we have seen, is truth in virtue of the semantic rules; and one of the semantic rules specifies the universe of discourse. (Frost-Arnold 2013: 77)

What has been more controversial, however, is whether this way of dealing with Gödel's incompleteness results is not just unusual but also misguided. There has certainly been no shortage of critics, one who is particularly blunt is John von Neumann:

> [...] I regard Carnap's things as naive and feeble. Carnap simply does not possess the knowledge minimally required to address such matters – let alone to say something new. [...] I am specifically annoyed that, while Gödel's name is constantly on Carnap's lips, it is obvious that he absolutely did not understand the real meaning of Gödel's results. (von Neumann 2005: 203)

The dialectical situation is very intricate, however, especially given the deflationary nature of Carnap's position I discussed in the previous chapter. If he were a traditional conventionalist who wanted to account for mathematical truth in a non-circular way – i.e. without already presupposing any mathematical notions in the explanation – then it would be easy to object to Carnap's use of infinite sets of numerals or intended models. But since this is clearly not his ambition, it is harder to pin down what, if anything, is wrong with his liberal conception of linguistic rules. In the course of this thesis I will argue that there in fact *is* a problem, but it will take a while to reach that point. We are, however, in a position to reconsider the question of what Carnapian frameworks are supposed to be, which will also shed light on the issue of whether any traces of conventionalism remain in Carnap's position.

Contrary to what I assumed for expository purposes, a Carnapian framework is not *one* formal theory (understood as a set of axioms and inference rules), but rather a *package* of such theories: in particular a framework consists of an object language *and* a metalanguage, or even a hierarchy of metalanguages (Leitgeb and Carus 2020: section 1.2). So while the theory of Peano arithmetic may well be the object language of a framework for doing mathematics, we should better not call PA *itself* the framework. This is because in many frameworks the metalanguage will play an essential role. The definition of analyticity which secures the truth of Con_{PA}, for instance, is stated in a metalanguage, and *Logical Syntax* in fact contains a proof that no sufficiently powerful and consistent object language can contain its own analyticity-predicate, which is a version of Tarski's argument for the indefinability of truth (Carnap 1937a: 219).[7] So we should understand the framework of mathematics as the combination of PA and a particular metalanguage.

Earlier I pointed out that when Carnap talks about mathematical truth being determined by, or flowing from, linguistic rules, this should not be read as a metaphysical thesis. I did not say anything positive about what Carnap *does* mean by such claims then, but we are now in a position to do so. One can in fact distinguish two senses of determination here. Sentences like "2+2=4" and "there are infinitely many primes" can be syntactically derived from the Peano axioms, and so there is one sense of determination that coincides with derivability. In the terminology Carnap uses, we can say that the truth of many mathematical sentences is determined by the *rules of derivation* (d-rules) of a framework (Carnap 1937a: 99).

Sentences like Con_{PA}, on the other hand, do not follow from the d-rules, but rather from what Carnap calls *rules of consequence* or c-rules (Carnap 1937a: 100). The definition of analyticity we saw earlier is an example of such a c-rule. It is important to note that the relevant rule was essentially *infinitary* and *non-effective*, since the analyticity of an all-quantified claim is established on the basis of the analyticity of infinitely many other sentences. This is then the sense in which Carnap overcomes Gödel's incompleteness result by relaxing the requirement that theories need to be recursively presented: since on his conception linguistic frameworks can include non-effective infinitary rules, there is no way to recursively represent a relation such as 'analytic in the framework of mathematics'.[8]

Carnap's approach has struck many philosophers as outlandish, however, including Gödel himself. In a paper to which we will turn soon, Gödel writes that for mathematics to be determined by linguistic rules in any meaningful sense, the relevant rules need to use "only finitary concepts referring to finite com-

[7] Based on the historical evidence available it is not easy to tell whether Carnap developed this argument independently of Tarski. For a timeline see Woleński 2005.

[8] One feature of this view that will be very important later is that there can be frameworks with identical d- but different c-rules.

binations of symbols", and maintains that this requirement "should be beyond dispute" (Gödel 1995a: 341). A well-known example of an inference that fails to meet this requirement is the so-called ω-rule, which allows one to infer a universal generalisation based on infinitely many premises:

ω-Rule

$$\frac{\phi(0), \phi(1), \phi(2), ...}{\forall x \phi(x)}$$

In *Logical Syntax* Carnap explicitly discusses this rule, and his attitude is diametrically opposed to that of Gödel:

> Tarski discusses [... the ω-rule] and rightly attributes to it an "infinitist character". In his opinion: "it cannot easily be harmonized with the interpretation of the deductive method that has been accepted up to the present"; and this is so far as this rule differs fundamentally from the [... finitary rules] which have hitherto been exclusively used. In my opinion however, there is nothing to prevent the practical application of such a rule. (Carnap 1937a: 173)

Adjudicating this disagreement will be a complex matter, and I will soon move on to discuss various objections to Carnap's philosophy of mathematics, starting with Gödel's own. It should be stressed though that, on its own terms, the position Carnap recommends seems at least *coherent*. His aim was to argue that there is a linguistic framework in which each purely mathematical sentence is either analytic or synthetic, including undecidable ones such as Con_{PA}. And given his liberal conception of linguistic rules, he seems to have succeeded. The claim that Con_{PA} is analytic in a mathematical framework requires that in the metalanguage M of this framework Con_{PA} can be established – and Carnap has presented a metalanguage which achieves this, even though with the help of a non-recursive rule.

Before moving on to objections we will now briefly look into the history and development of Carnap's position, for this will elucidate how he came to give up the requirement that all legitimate theoretical notions must be decidable.

2.1.3 Towards Internalism

It is possible to construe the issues raised by Carnap's response to Gödel's incompleteness results as a debate about how quantification works. One important upshot of Gödel's theorems is that no recursive first-order theory is such that it only has the standard model of arithmetic, or models isomorphic to it, as its domain. This raises the question whether we are justified to hold that *despite* this underdetermination the quantifiers of our preferred arithmetical theory range only over the standard numbers, which can in turn explain why we regard some mathematical sentences as true even though they are independent of the axioms. Initially it might seem that anyone who answers this question in the affirmative must be a Model-in-the-Sky Platonist, who thinks that the world provides a distinguished domain of quantification of mathematical objects. But according to Carnap's internal Platonism we can have a determinate domain of quantification without buying into the metaphysics: to say that we quantify over the standard numbers is no metaphysical posit, but just a claim made in the metalanguage.

Carnap's acceptance of this position seems to have been the result of an exchange with Gödel about second-order quantification and impredicative definitions. This had been a hot topic in the philosophy of mathematics in the 1920s, but we only need a very brief introduction for our purposes. A definition is called impredicative if "it generalizes over a totality to which the entity being defined belongs" (Linnebo 2020b). Consider the following example:

> Let *n* be the least natural number such that *n* cannot be written as the sum of at most four cubes. (Linnebo 2020b).

The entity to which this definition applies is a particular natural number, and since the definiendum includes quantification over natural numbers we are dealing with an impredicative definition. Analogously, a definition that characterises a certain *property* is called impredicative if its definiendum contains quantification over all properties.

To many philosophers impredicative definitions had seemed suspicious, at least when used in mathematics. But, as Ramsey famously observed, it is not so clear why, since in the realm of concrete objects they seem perfectly in order:

> [...] we may refer to a man as the tallest in a group, thus identifying him by means of a totality of which he is himself a member without there being any vicious circle. (Ramsey 1925: 368)

It seems that the ontological status of the entities or properties that are being quantified over matters. As Gödel sums it up in retrospect, if one accepts a constructivist view of mathematics, according to which definitions not merely *describe* entities but rather bring them into existence in the first place, then

> [...] there must clearly exist a definition (namely the description of the construction) which does not refer to a totality to which the object defined belongs, because the construction of a thing can certainly not be based on a totality of things to which the thing to be constructed itself belongs. If, however, it is a question of objects that exist independently of our constructions, there is nothing in the least absurd in the existence of totalities containing members, which can be described (i.e., uniquely characterized) only by reference to this totality[.] (Gödel 1983: 456)

In his "Foundations of Mathematics" Ramsey assumed a *realist* conception of properties. Carnap was familiar with this debate, and in a paper from 1931 comments as follows:

> [...] I think we should not let ourselves be seduced by it into accepting Ramsey's basic premise; viz., that the totality of properties already exists before their characterization by definition. Such a conception, I believe, is not far removed from a belief in a platonic realm of ideas which exist in themselves, independently of *if* and *how* finite human beings are able to think them. [... It seems to me that] we should call Ramsey's mathematics "theological mathematics", for when he speaks of the totality of properties he elevates himself above the actually knowable and definable and in certain respects reasons from the standpoint of an infinite mind which is not bound by the wretched necessity of building every structure step by step. (Carnap 1983: 50)

This passage is intriguing, since here Carnap's position appears to be perfectly in line with what was, and mostly still is, widely accepted among philosophers of mathematics: namely that the price of giving up constructivism is a certain degree of Platonism, which may then be questioned on metaphysical grounds. Within a few years, however, Carnap's attitude changed radically, so let us see how that came about.

Carnap had sent an early draft of his *Logical Syntax* to Gödel, which contained a substitutional definition of second-order quantification. Gödel identified a problem with this approach: for Carnap's definition to work correctly, every use of higher-order quantification must ultimately be reducible to a formula which does not itself contain higher-order quantifiers. As Gödel pointed out, however, Carnap's definition does not actually deliver this result (Gödel 2003a: 347). Consider the following formula:

(1) $\forall P\ P(0)$.

If we use substitutional second-order quantification, the truth of (1) depends on the truth of all substitution instances of (1). Gödel now draws attention to the fact that since '$\forall F(F(x))$' is a predicate definable in the language of (1), it can be substituted for P, and hence the following is a substitution instance of (1):

(2) $\forall F\ F(0)$

But (2) is obviously just a notational variant of (1) in which second-order quantification is still present, and a repeated application of Carnap's definitions would not make any difference. His method thus fails to eliminate second-order quantification in the way that is required.[9]

To solve this problem Gödel recommends the following solution:

> In my judgment, this error may only be avoided by regarding the domain of the function variables not as the predicates of a definite language, but rather as all sets and relations whatever. (Gödel, Letter to Carnap, 11 September 1932 (Gödel 2003a: 347))

The suggestion is thus to read the second-order quantifier *objectually*, and in such a way that it also ranges over properties that are not definable within the relevant language. This seems to be an endorsement of Ramsey-style realism about properties, however, and it is hence understandable that Carnap's first reaction to this proposal is a certain amount of hesitancy and confusion:

> You say: it [= the universal quantifier] must range over "all sets"; but what does that mean? (Carnap, Letter to Gödel, 25 September 1932 (Gödel 2003a: 351))

While working through Gödel's suggestion, Carnap eventually realises that he needs a semantics in which the second-order quantifiers range over *valuations*. As he explains in later work, valuations are sets of numerals (or natural numbers):

[9] See Goldfarb 2003: 338 and Flocke 2019 for discussion.

> By a possible *valuation* (syntactical designation, $\mathfrak{B}$) for '*F*' (i.e. a value assigned to '*F*') we shall here understand a class (that is to say, a syntactical property) of accented expressions. (Carnap 1937a: 107)[10]

Carnap is still concerned about the fact that while there are *un*countably many valuations, in each particular language only *countably* many of them will be describable, and expresses his reservations as follows:

> It seems to me not questionable if "analytic in the language S" cannot be defined in a semantics that is formalised in S, but only in a semantics that is formalised in a more extended language S_2. But to operate with a concept for which there is no language at all in which it can be rigorously defined is certainly rather questionable. (Carnap, Letter to Gödel, 25 September 1932 (Gödel 2003a: 351))

Only two days later, however, Carnap writes another letter to Gödel, in which he announces that he has overcome the problem:

> Yesterday I found the solution: The locution "for every valuation ..." that occurs in the definition can still be expressed in a semantics formulated in a definite language, namely by "$[F](...)$", since a valuation is of course a semantic predicate. This is possible even though in the semantics under consideration not all possible valuations, that is, predicates, can be defined. (Carnap, Letter to Gödel, 27 September 1932 (Gödel 2003a: 355))

What is the crucial insight Carnap had between these two letters? Apparently it was that all we need to do in order to give a satisfactory semantics of second-order quantification is to have a language in which we can quantify over all valuations. And since valuations are just *sets of numbers*, this only requires that we can use a language in which (first-order) quantification over those kinds of entities is possible, which Carnap here seems to presuppose. The indefinability of some of the valuations would be a problem if we could only quantify over things we can pick out using a description, an assumption Carnap now rejects.

It is interesting to note that in the first version of *Logical Syntax* that Gödel criticised, Carnap still seemed to have a quite different conception of what a semantics for second-order language is supposed to achieve. The idea behind the

10 See Flocke 2019: 398f for more on Carnap's terminology.

substitutional semantics given was presumably to *eliminate* second-order quantifiers by way of a decidable procedure of reduction. The new approach clearly doesn't deliver as much, as Carnap is now happy to leave second-order quantifiers unreduced, and only demands that there is a clear specification of their truth-conditions. In the version of *Logical Syntax* that appeared in print, Carnap describes his considered approach as follows:

> [...] the definition must not be limited to the syntactical properties which are definable in S, but must refer to all syntactical properties whatsoever. [... The question then becomes]: can the phrase "for all properties ..." (interpreted as "for all properties whatsoever", and not "for all properties which are definable in S") be formulated in the symbolic syntax-language S? This question may be answered in the affirmative. The formulation is effected by the help of a universal operator with a variable 𝔭, i.e. by means of '(*F*)(...)', for example. (Carnap 1937a: 113f)

Furthermore Carnap refers back to his earlier criticism of Ramsey, and stresses that his account of second-order quantification is strictly non-metaphysical:

> But do we not by this means arrive at a Platonic absolutism of ideas, that is, at the conception that the totality of all properties, which is non-denumerable and therefore can never be exhausted by definitions, is something which subsists in itself, independent of all construction and definition? From our point of view, this metaphysical conception – as it is maintained by Ramsey for instance (see Carnap [*Logizismus*] p. 102) – is definitely excluded. We have here absolutely nothing to do with the metaphysical question as to whether properties exist in themselves or whether they are created by definition. (Carnap 1937a: 114)

There is something puzzling about this attitude, however. For Carnap explains second-order quantifiers in terms of numbers and sets, where some of the relevant sets (valuations) will of necessity be indefinable. And, one might reasonably think, isn't this just as problematic as postulating a realm of indefinable properties? Carnap's response to this would presumably be to re-apply his internalist treatment to the language in which the definition of second-order quantifiers is given, and hence make these worries disappears. In the end the viability of his treatment of second-order quantifiers thus depends on his treatment of arithmetic and set theory.

At the moment I want to stress another point though. It is striking that in one of his letters Gödel himself seems to *agree* that the strategy he suggests to Carnap is distinct from, and hence an alternative to, mathematical Platonism:

> This doesn't necessarily involve a Platonistic standpoint, for I assert only that this definition (for ″analytic″) be carried out within a definite language in which one already has the concepts ″set″ and ″relation″. (Gödel, Letter to Carnap, 11 September 1932 (Gödel 2003a: 347))

Given that Gödel is generally known as *the* archetypical Platonist, this raises many questions. Did Gödel actually think that Carnap's account in *Logical Syntax* was viable at the time, and only changed his mind about this later? If so, was he even opposed to Platonism itself in the 1930s, and only came to endorse it later on? I think it would be a mistake to make too much of Gödel's remark, however, even though Carnap was presumably very pleased to hear it.[11] For it seems to me that some of the agreement between Carnap and Gödel in the letters is merely apparent. In the same letter in which Gödel writes that ″you have understood my suggestions about the definition of 'analytic' entirely as I meant them″, he also makes the following comment:

> I believe moreover that the interest of this definition does not lie in a clarification of the concept ″analytic″, since one employs in it the concepts ″arbitrary sets″, etc., which are just as problematic. (Gödel, Letter to Carnap, 28 November 1932 (Gödel 2003a: 357))

This sounds like a version of the objection I hinted at above, according to which nothing is won by moving from indefinable properties to indefinable sets. And I therefore think that at a fundamental point, Carnap and Gödel's positions are opposed to each other already in 1932. Carnap wants to use the concept of analyticity in order to elucidate mathematical truth, whereas Gödel thinks that,

[11] Of course it is hardly obvious what exactly Gödel's own Platonism amounts too, see for instance Potter 2001. Goldfarb provides some evidence to support the view that Gödel was a less committed Platonist in the 1930s than he later became (Goldfarb 2005: 191fn5), but Gödel himself maintained that he was already a Platonist during the Vienna period (Awodey and Carus 2010: 260fn13).

if this notion is of any interest at all, it definitely cannot play the theoretical role Carnap has in mind.[12]

While this is somewhat speculative, it is uncontroversial that by the 1950s Gödel explicitly rejected Carnap's position of *Logical Syntax*. We will now move on to the second part of this chapter in which Gödel's later arguments, and some Carnapian rejoinders, are discussed.[13]

2.2 Gödel's Consistency Argument

2.2.1 Is Mathematics Syntax of Language?

In 1953, Paul Arthur Schilpp asked Gödel for a contribution to the volume on Carnap's philosophy, and suggested the title "Carnap and the Ontology of Mathematics". In response, Gödel wrote that he wouldn't be able to produce of paper of such broad scope, but agreed to make a smaller contribution for which he gave the provisional title "Some Observations on the Nominalistic View of the Nature of Mathematics". Over the next few years he went on to produce six drafts of this paper which all have the title "Is Mathematics Syntax of Language?". In 1959, Gödel then let Schilpp know that he won't be able to deliver the promised paper after all. Gödel gives the following reason for his dissatisfaction with what he had written so far:

> It is easy to allege very weighty and striking arguments in favor of my views, but a complete elucidation of the situation turned out to be more difficult than I had anticipated, doubtless in consequence of the fact that the subject matter is closely related to, and in part identical with, one of the basic problems of philosophy, namely the question of the objective reality of concepts and their relations. (Gödel, Letter to Schilpp, February 3rd 1959 (Gödel 2003b: 244))

[12] Goldfarb interprets this letter in a different way: he reads Gödel as having reservations about the notions of analyticity and arbitrary sets themselves, and hence writes that "[t]his may be the only time that Gödel expresses a position that is less ontologically exuberant than Carnap's" (Goldfarb 2009: 120fn5). On my reading Gödel has no qualms with the notion of arbitrary sets as such, but merely points out that the notion of analyticity is of no help in understanding such mathematical concepts.

[13] The relevance of Carnap's account of impredicative definitions to his views on ontology has recently been debated in the secondary literature. Gregory Lavers has argued that Carnap's approach does not really avoid Platonism, since it "only pushes the problem back a step" (Lavers 2015: 274f), and refers to Beth's much later paper. In response, Vera Flocke has argued that Lavers' objection is not the same as Beth's (Flocke 2019: 397fn86). As I will show in section 4.2.2 there is indeed a connection between Beth's point and questions of ontology, but it is not obvious whether my reconstruction matches what Lavers had in mind.

Four of Gödel's drafts were published in the 1990s, and they have generated a good amount of discussion. The drafts are of differing length, and some contain more than a single argument. In the following I will focus on a line of reasoning that can be found in all versions, and has also been most widely commented upon in the secondary literature: namely the *argument from consistency*, which purports to show that the second incompleteness theorem refutes the project of *Logical Syntax*. In this part of the chapter I will present the argument, introduce some responses that have been given on Carnap's behalf, and then evaluate how forceful Gödel's considerations are in light of these rejoinders.

Another topic that is prominent in some of Gödel's drafts, but which I will not focus on here, concerns the *application* of mathematics. Briefly put, one point Gödel makes is that Carnap's claim that mathematics has no empirical content is false or at least misguided, since often mathematical claims *do* have empirical consequences:

> [...] on the grounds of a proof for Goldbach's conjecture (which says that every even number is the sum of two primes), it can be predicted that a computing machine which is empirically known to work reliably will find two primes whose sum is some given large number N. (Gödel 1995a: 339)

This suggests that even if Carnap *could* maintain that all of mathematics is analytic, he *should* refrain from doing so, since this way of setting things up doesn't actually capture how we normally construe the relationship between mathematics and the empirical world. This is certainly an important consideration, and Gödel's line of reasoning sometimes resembles the more well-known arguments for holism by Quine.[14] I will focus on the part of Gödel's argument that concerns the question of consistency, however, since this will naturally lead us to E. W. Beth's related but distinct – and, so I will argue, more successful – argument in chapter 4.

What Gödel wants to refute is the thesis that *mathematics is syntax of language*, which he thinks can be found in Carnap's *Logical Syntax*. He characterises it as follows:

> I. Mathematical intuition, for all scientifically relevant purposes, in particular for drawing the conclusions as to observable facts occurring in applied mathematics, can be replaced by conventions about the use of symbols and their application.

[14] See Lavers 2019 for more discussion.

> II. In contradistinction to the other sciences, which describe certain objects and facts, there do not exist any mathematical objects or facts. Mathematical propositions, because they are nothing but consequences of conventions about the use of symbols and, therefore, are compatible with all possible experiences, are void of content.
>
> III. The conception of mathematics as a system of conventions makes the a priori validity of mathematics compatible with strict empiricism. For we know a priori, without having to appeal to any a priori intuition, that conventions about the use of symbols cannot be disproved by experience. (Gödel 1995a: 356)

It is doubtful whether Carnap himself would have accepted this characterisation of his aims in *Logical Syntax* without any revisions. Since Gödel did not submit any draft in the end, Carnap apparently never got so see the argument, and so we can only speculate how he would have reacted. While one might quibble with some of Gödel's formulations, I-III are nevertheless not completely off the mark or uncharitable. For it is clearly true that Carnap thinks we neither need to postulate a faculty of mathematical intuition nor mathematical objects, and also holds that his account makes mathematics acceptable for empiricists.

The case of mathematical objects is especially interesting, since Gödel seems to believe that Carnap has changed his mind about this point since the 1930s:

> From what he says in [... *Empiricism, Semantics, and Ontology*] it follows that at present he does not object to associating, in scientific semantics, mathematical objects to formulae as their denotation. However he maintains that the philosophical question about the objective existence of mathematical objects does not refer to this "internal" existence, but means whether these objects formally introduced by axioms "really" exist. An answer to this question is asserted to have no "cognitive content", i.e., the question is considered to be meaningless, while formerly it was answered negatively; or else Carnap has changed his opinion about internal existence in mathematics. (Gödel 1995a: 355)

I think that Gödel here overestimates the impact of Carnap's turn to semantics, however. It is true that during the time of *Logical Syntax*, Carnap did not explicitly have something like the internal/external-distinction, and would also not have asserted that numerals refer to numbers, since he rejected the notion of reference outright. On the other hand, even during the syntactic period he could already have said that there is an unproblematic sense in which it is true that numbers exist, since sentences like "$\exists x(x = 1)$" are true in language II of *Logical Syntax*. This seems to correspond to the internal existence statements of ESO, and when pressed whether there is not more to the existence of numbers than is captured by statements of his kind, Carnap would likely have dismissed this worry as meaningless and confused as well. There is thus much more continuity between Carnap's earlier and later views than Gödel makes it out to be.

This means, on the other hand, that Gödel's argument against *Logical Syntax* also applies to ESO, and so let us turn to it now. For a concise presentation it will be useful to introduce the notion of a *conservative extension*:

> Theory T^* is a conservative extension of a theory T iff
>
> (i) every theorem of T is a also theorem of T^* and
>
> (ii) every theorem of T^* that is expressed in the language of T is also a theorem of T.

In order to appreciate Gödel's consistency argument, it is best to imagine that we accept some theory of the world – I call this the *base theory* – which does not already contain mathematics. Suppose we now want to extend the base theory by a mathematical theory. The argument against the idea that mathematics is syntax of language now runs as follows:

(1) If some theory is adopted as a convention, it must be known that it is a conservative extension of the base theory.

(2) A theory that extends a consistent base theory is conservative only if the former theory is consistent.

(3) So: A mathematical theory can be adopted as a convention only if it is known whether this theory is consistent.

(4) We know from Gödel's incompleteness theorems that for any sufficiently strong mathematical theory, we need even stronger mathematics to prove that theory's consistency.

(5) So we need to rely on mathematical intuition at some point in order to know that the theory we want to adopt as a convention is consistent.

(2) is obviously true if we accept the principle of explosion, as Gödel and Carnap both do. The step from (4) to (5) relies on several background assumptions, which I will briefly discuss and then aside, in order to focus on the most controversial premise (1) for the rest of the chapter. Gödel's thinking seems to be that the only way to gain *knowledge* of consistency is through a consistency *proof*, or through somehow recognising that the axioms of a mathematical theory are *true*, which entails that they are consistent. The latter act of grasping the truth of axioms is what he calls mathematical intuition. Gödel does consider a third option, namely coming to know the consistency of a system through empirical observation, and comments on this as follows:

> [...] it may be argued that, although transfinite mathematical axioms clearly must not be used, it is permissible to use empirical induction. E.g., consistency might be based on the fact that no contradiction has arisen so far. Now it is true that, if consistency is interpreted to refer to the handling of physical symbols, it is empirically verifiable like a law of nature. However, if this empirical consistency is used, mathematical axioms and sentences completely lose their "conventional" character, their "voidness of content" and their "apriority" [...] and rather become expressions of empirical facts. (Gödel 1995a: 342)

This seems a reasonable point in general, and is definitely justified when it comes to Carnap. For we have already seen that one consequence of treating mathematics as analytic is that empirical evidence – which for Carnap is expressed by means of protocol sentences – can neither support nor undermine mathematical sentences. Defenders of Carnap have therefore focussed their attention on criticising Gödel's premise (1), so let us therefore move on to scrutinising it.

2.2.2 Content and Knowledge

Gödel thinks that we should accept the following claim:

(1) If some theory is adopted as a convention, it must be known that it is a conservative extension of the base theory.

His own explanation for why (1) is plausible goes as follows:

> Moreover a rule about the truth of sentences can be called syntactical only if it is clear from its formulation, or if it somehow can be known

> beforehand, that it does not imply the truth or falsehood of any "factual" sentence (i.e., one whose truth, owing to the semantical rules of the language, depends on extralinguistic facts). [...] The requirement under discussion implies that the rules of syntax must be demonstrably consistent, since from an inconsistency every proposition follows, all factual propositions included. (Gödel 1995a: 339)

To illustrate this, assume we have a non-mathematical and consistent base theory in which we can talk about physical objects and their properties. Suppose we now extend this theory by adding mathematics. If the mathematical theory we add happens to be inconsistent, then the whole theory becomes inconsistent as well, and hence we can suddenly derive various empirical claims about physical objects, some which we would regard as false: "there are seventeen blue tables in room R45", for instance. But this, so Gödel, shows that the acceptance of the mathematical theory cannot have been a matter of convention after all, since he regards it as definitional of a convention that its adoption cannot imply any empirical predictions. This is because such predictions would make the convention *refutable*, and hence the distinction between a hypothesis and a convention is blurred. As a consequence of this, so Gödel, a mathematical theory can only be adopted as a convention if it is known to be consistent.

Scholars of Carnap have pushed back against this argument. Since the 1990s Goldfarb and Ricketts have variously argued that for Carnap it is not in fact true that "from an inconsistency every proposition follows, all factual propositions included", an assumption without which Gödel's argument loses its force (Goldfarb and Ricketts 1992, Goldfarb 1995). By this Goldfarb and Ricketts don't mean to say that Carnap accepts some form of paraconsistent logic in which the principle of explosion fails, however. So what do they have in mind? To understand their response it is useful to consider what Carnap means by the *empirical content* of a sentence:

> The content of a sentence S is the set of non-valid sentences which follows from S. (adapted from Carnap 1937a: 175 (§49)).

It is a consequence of this conception that in an inconsistent framework, in which *every* sentence is valid, *no* sentence has any empirical content. And, so Goldfarb and Ricketts stress, this means that an inconsistent framework really makes *no* claims about the empirical world at all, contrary to what Gödel supposes. Consider again the consistent base theory to talk about physical objects, and suppose that the sentence "there are seventeen blue tables in room R45" is false in it. It is of course the case that once we extend the base theory by an inconsistent mathematical theory, the *sentence* "there are seventeen blue tables in room R45" will now be derivable. But this sentence, so to speak, does not express the same *empirical proposition* which it does express in the consistent base theory. In the inconsistent theory, every sentence can rather be said to express one and the same trivial proposition. Rightly understood, the adoption of an inconsistent mathematical theory does therefore not result in any false empirical predictions after all.

One might express this by saying that for Carnap the notion of an empirical fact only makes sense *within* a particular framework, as Goldfarb does here:

> [...] applied to *Logical Syntax*, Gödel's argument would presuppose a notion of empirical fact that transcends or cuts across different linguistic frameworks. However, as the Principle of Tolerance strongly suggests, it is central to the view of *Logical Syntax* that any such language-transcendent notion be rejected. Rather, the notion of empirical fact is given *by way of* the distinction between what follows from the rules of a particular language and what does not. Thus, on this reading Carnap's view undercuts the very formulation of Gödel's argument. (Goldfarb 1995: 328)

While Gödel worried that an inconsistent theory would make various *false* claims about the empirical world, the real issue with inconsistent theories is rather that they cannot make *any* claims with empirical content at all.

In itself this is an effective reply, for it does seem true that Gödel presupposes that sentences express certain propositions in a way that is not directly mediated by the rules of the theory the sentences are a part of.[15] But one can question whether this move is really a *defence* of Carnap's overall philosophy, or rather points towards an even more fundamental problem. Michael Potter illustrates this concern with the following scenario:

> Imagine the case of a language which has been used for many years with apparent success despite the existence within its mathematical part of an abstruse and as yet undiscovered contradiction (the Burali-Forti paradox, say). We seem to be forced by Carnap's account to say that despite appearances this language is not succeeding in saying anything about the world. (Potter 2000: 276)

[15] The relationship between sentences and propositions also plays an important role in contemporary discussions of conventionalism, see for instance Warren 2015b.

Generalising from this, Potter argues as follows: it seems to be an undeniable fact that our language contains at least *some* sentences that make empirical claims about the world, as an example take "this table is black". On Carnap's conception only consistent languages can say anything about the world. Hence in order to know that "this table is black" makes an empirical claim, I need to know that our language is consistent. But if we suppose that our language contains mathematics, this knowledge is hard to come by – we have at most inductive evidence for consistency, but no certainty. This leads to the result that we only have defeasible reasons to think that "this table is black" makes a claim about the empirical world at all. Carnap is therefore committed to a radical form of holism about empirical content, which Potter regards as absurd:

> Carnap's view makes it an experimental fact that I have a conception of an empirical world at all. That is as close to a straightforward contradiction as one is likely to encounter in philosophy. (Potter 2000: 277)

It therefore seems that what Goldfarb and Ricketts intended to be a defence of Carnap's position, namely the fact that the notion of empirical content is characterised in a framework-dependent and holistic way, has been turned into an *improved* version of Gödel's consistency argument. Once again we have arrived at the conclusion that Carnap is in trouble because, due to the second incompleteness theorem, we cannot have the secure knowledge of consistency one would need for Carnap's position to be viable.

Potter's improved version of Gödel's argument has in turn been criticised by Steve Awodey and André Carus, and I will discuss their reaction to it in the next section. They also propose another response to Gödel, however, since they speculate that the strategy pursued by Goldfarb and Ricketts is very different from what Carnap himself would have said. Awodey and Carus's suggested response is to maintain that where Gödel demands *known* consistency, Carnap only needs *de facto* consistency:

> In short, where Gödel says "the rules of syntax must be **demonstrably** consistent, since from an inconsistency *every* proposition follows" (ibid.), he should correctly say "the rules of syntax must **be** consistent, since from an inconsistency *every* proposition follows". (Awodey and Carus 2004: 208, my emphasis in bold)

Applying this strategy to my earlier reconstruction of Gödel's argument yields the following replacement for premise (1):

(1*) If some theory is adopted as a convention, **it must be** a conservative extension of the base theory.

Revised in this way, Gödel's argument is no longer a challenge for Carnap. Awodey and Carus assume that many mathematical theories are in fact consistent, even if we cannot prove them to be so in a non-circular way. Given (1*) there is hence no obstacle to adopting them as a convention.

Does this response suffice as a defence of Carnap? I think not. For one thing, the assumption that there are determinate facts about consistency, and matters of syntax more broadly, is not as innocent as Awodey and Carus seem to assume. It is certainly plausible that Carnap *needs* such facts, for otherwise a non-epistemological version of Potter's argument could be given: if there is no fact of the matter whether our language is consistent, then there is no fact of the matter whether we talk about an empirical world either, which seems to be an even more bizarre consequence. As I will argue in section 4.2.3, however, accounting for facts about syntax is actually very hard for an internalist like Carnap.

Let us provisionally take syntactic facts for granted though. Even then Awodey and Carus' reply only suffices to deal with Gödel's argument as it was originally presented, while the objection Potter raises nevertheless remains salient. Carnap still seems stuck with a conception of language according to which we can never be certain that we manage to talk about the empirical world at all, and further arguments are needed to dispel the worry that this is an unpalatable view. Awodey and Carus are aware of this, and I will discuss their response to Potter in the next section.

2.2.3 Frameworks and Natural Language

Awodey and Carus's reply to Potter has two parts. The first part seems to misconstrue the nature of Potter's argument, however, and so I will set it aside relatively quickly. The second part of their reply is more interesting, but since it concerns core assumptions about how linguistic frameworks relate to natural language it will not be possible to conclusively evaluate whether it is a successful defence of Carnap right here. This topic will occupy us throughout the thesis, however, and hence introducing it at this point is useful.

The first part of Awodey and Carus's reply to Potter attacks the idea that there is such a thing as an indefeasibly descriptive sentence:

> It is fundamental to the later Carnap's view [...] that there is *no* fixed

> partition, antecedent to any language, of sentences into analytic ones and synthetic ones. *Any* sentence whatever, including Potter's favorite "This table is black", could, if it were for some reason convenient, be made into a constitutive language rule, and thereby deprived of its descriptive capacity within that language. No sentence, regarded in isolation, is *inherently* descriptive. (Awodey and Carus 2004: 212f)

What is again stressed here is that for Carnap, each sentence can be considered as meaningful only in the context of a framework. Whether a particular sentence counts as descriptive is settled by the framework rules, and is not an intrinsic property of that sentence in isolation. Furthermore, Awodey and Carus allude to Carnap's idea that we could even construct frameworks in which synthetic sentences about observations are linguistic rules. Carnap calls such rules P-rules, and primarily envisaged laws of nature to play this role. This suggestion has led to some confusion, since it might seem that if we can construe laws of nature as framework rules, this amounts to stipulating that certain laws of nature actually hold – which is obviously not within our power (Coffa 1991: 351f). Rightly understood the role of P-rules is much more mundane, however. In practice the only difference between using a framework with P-rules and using one without is that we might have to switch between frameworks more frequently:

> Relative to such a language, these "P-rules" as he called them [...], could not be falsified (if we had some reason to doubt them, we would have to change the language). (Awodey and Carus 2004: 206fn10)

The question now is, however, whether any of this undermines Potter's argument. I think not, for it seems to me that Awodey and Carus commit a kind of quantifier shift fallacy, and thus misinterpret the role of Potter's example sentence "this table is black". Contrary to what they suggest, the idea was presumably not that *this* particular sentence is descriptive in *every* framework. If this were the claim, then Awodey and Carus's objection to intrinsically descriptive sentences would have force. But I take it that Potter's thought was rather that *for every* framework which is sufficiently rich to purport to make sense of the empirical world – i.e. setting aside purely mathematical frameworks – there is *some* sentence which every user of this framework will want to classify as descriptive.

Applied to our own language, this means that it does not really matter whether "this table is black" is such a sentence, it was just a convenient and plausible example. All that Potter really needs is the following assumption:

> There is at least one sentence of our language that uncontroversially describes the empirical world.

And this is a weak claim which, I think, everyone should agree to, and which is also not undermined by the observation that *being descriptive* is not an intrinsic but rather a framework-relative notion. I therefore conclude that the first part of Awodey and Carus's reply fails to get at the core of the improved consistency argument.

The second part of their reply is, on the one hand, much more promising, but, on the other hand, also harder to evaluate. One fundamental assumption we have been making is that the things Carnap says about linguistic frameworks – in particular that inconsistent frameworks have no empirical content – also apply to natural language, i.e. the language we are speaking. Without this assumption the charge that, according to Carnap, we have at most defeasible inductive evidence for thinking that we are talking about an empirical world would not stick. And it is this presupposition that Awodey and Carus go on to attack:

> Though his conception of the scientific, theoretical language was holistic [...], he specifically saw the practical, intuitive part of language as distinct from it and serving a different purpose. The ordinary linguistic competence of human language user communities is the practical setting (*not* a precise framework) within which theoretical languages are precisely specified [...]. The descriptive capacity of a language is not destroyed by the discovery of an inconsistency since we can change the rules or postulates of the theoretical language without also changing the vocabulary of ordinary objects in our our everyday language. The purely descriptive sentences in the latter are unaffected by an inconsistency in the theoretical language. (Awodey and Carus 2004: 213)

Their suggestion seems to be that since ordinary language is not a precise linguistic framework, questions of consistency and empirical content cannot even arise in the way they do for formalised frameworks. And in this way the apparently absurd conclusion that we can only know that we're talking about anything empirical at all if we know that our language is consistent can be averted. Even though we can never be completely sure that some sophisticated formal language manages to be descriptive, we can at least rest content with our intuitive ordinary language – a move that brings Quine's *acquiescence in one's home* *language* to mind (Quine 1969b: 49).[16]

If successful, the *natural language is different* reply would indeed make the problem posed by Potter much less pressing. In order to properly evaluate it we need to know more about how exactly Carnap conceives of the role of natural language and its relation to linguistic frameworks, however, and we will look into this matter in chapter 4. One can already note two reasons to be skeptical about this reply though, one exegetical and the other systematic.

On the exegetical side, the worry is that the reply assumes that there is some essential difference between natural and formal languages that makes the former immune to Potter's objection, whereas Carnap himself seems to deny that there is such a difference. This emerges quite clearly in an exchange between Carnap and Strawson on the point of explicating natural language concepts. Carnap introduces a position called *linguistic naturalism*, which he describes as follows:

> In my view, a language, whether natural or artificial, is an instrument that may be replaced or modified according to our needs, like any other instrument. *For the naturalists, ordinary language seems to have an essentially fixed character and therefore to be basically indispensable,* just like our body with its organs, to which we may add accessories like eyeglasses, hearing aides, and the like, but which we cannot essentially change or replace. [...] Some naturalists seem to think that it is in principle impossible to learn an artificial language in any other way than by a translation into our mother tongue. (Carnap 1963b: 938, my emphasis)

Carnap then describes ways in which a group of speakers could learn a constructed language which do not involve translating the new language into some natural language. He sums up the point of this discussion as follows:

> My intention in making this point is not, of course, to propose the actual use of this method for learning a logical language, but merely to point out the theoretical possibility of such a procedure, and thereby to *refute the widespread view that constructed languages are not autonomous, but essentially parasitic, based on natural languages.* (Carnap 1963b: 938, my emphasis)

[16] This connection is also drawn in Ricketts 2004: 199, albeit in a different context. Interestingly Carus distances himself from the parallel with Quine in his book (Carus 2007: 291), which I will discuss some more in section 5.2.3.

These remarks strike me as being incompatible with an *exceptionalist treatment* of natural language, according to which certain questions and problems that arise for linguistic frameworks do not apply to natural languages at all. There is thus a tension between Awodey and Carus's strategy and Carnap's self-proclaimed anti-exceptionalism.

On a systematic level there is something unsatisfactory about the proposal as well. For unless we hear some more about how Carnap conceives of natural language, and in particular its ability to make contact with the world, it will seem that natural language here acts as a *deus ex machina* that is conveniently wheeled in to solve a problem for Carnap, but about whose inner functioning we know too little to say whether this move is more than ad hoc. If natural language remains a black box, critics of Carnap are therefore unlikely to find Awodey and Carus's response convincing, and justifiably so.

What is the upshot of Gödel's argument from incompleteness then? Neither Gödel's original nor Potter's improved version can directly *refute* Carnap's philosophy of mathematics, since his defenders have developed some powerful counterarguments. The suspicion that there is some deep problem brought about by the incompleteness theorems remains, however, since Carnap is pushed towards positions about the relationship between language, frameworks, and the world that are hardly innocent and uncontroversial. In a way this confirms Gödel's own assessment of the dialectical situation from the letter to Schilpp: a "complete elucidation of the situation" is very difficult, since "the subject matter is closely related to, and in part identical with, one of the basic problems of philosophy, namely the question of the objective reality of concepts and their relations" (Gödel 2003b: 244).

Chapter 3

A Quinean Interlude

In this chapter I will briefly discuss what is probably the most famous challenge ever raised to Carnap's philosophy: Quine's arguments against the viability of a distinction between analytic and synthetic statements. The aim is not a comprehensive evaluation of Quine's position, about which I will say more in chapter 8. It will be helpful to get a sense of the shape Quine's arguments take, however, in order to better appreciate the similarities and differences to Beth's argument. Like in the previous discussion of Gödel's argument from consistency, Carnap seems to be in a position to defend himself by rejecting some of Quine's basic assumptions, and so once again the case against the analyticity of mathematics remains inconclusive.

3.1 The Carnap-Quine Debate

That Quine rejected the notion of analyticity, which for Carnap plays a major role, is very well known.[1] But what exactly Quine's reasons for this rejection were, and how compelling his arguments are, is less well understood. There has been a lot of scholarship on the Carnap-Quine debate in recent years, and there is a reasonably broad consensus on the following points:[2]

(1) Quine thought that for the notion of analyticity to be philosophically viable, we need to explain the application conditions of the predicate 'analytic' for natural languages in behaviouristic terms.

[1] In his later work Quine was in fact happy to grant that there is *some* notion of analyticity that is coherent and may even be useful (Quine 1990: 54f, Quine 1997: 45), but it doesn't play the same theoretical role as the kind of analyticity he had attacked earlier.

[2] For an overview see Creath 2007. Some of the most important recent papers are Creath 1991, Stein 1992, O'Grady 1999, George 2000 and Gregory 2003b.

(2) Quine thought that the task described in (1) cannot be achieved.

(3) Carnap was willing to agree with (2), but rejected the assumption (1), i.e. that without such a behaviourist explication the notion of analyticity is philosophically useless.

(4) This disagreement demonstrates that Carnap and Quine have quite different conceptions of the relationship between formal languages with explicit rules and natural language, which reflect different conceptions of the task of philosophy.

(4) naturally makes the dialectical situation very complex, and it becomes hard to declare a winner in this debate – instead it is tempting to conclude that both are right on their own terms. Before we try to adjudicate the dispute, let us look into (1)–(4) in some more detail though.

If one only reads "Two Dogmas of Empiricism" (Quine 1951) it will probably not be immediately obvious that the bone of contention is the possibility of giving behavioural criteria for the application of 'analytic'. In the paper Quine discards various proposed definitions (relying, for instance, on the notion of synonomy) of 'analytic' as uninformative, but those attempted definitions are not actually ones Carnap himself advocated.

As we have seen, Carnap defines analyticity for specific languages or linguistic frameworks with explicit rules, since this allows him to characterise the analytic sentences as those that follow from the rules. He stresses the importance of this in an unpublished response to Quine's criticisms:

> It follows from this clarification that the analytic-synthetic distinction can be drawn always and only with respect to a language system, i.e., a language organized according to explicitly formulated rules, not with respect to a historically given natural language. (Carnap 1990: 432)

In "Two Dogmas", Quine's objection to this approach seems to be as follows: Carnap only manages to define predicates like 'analytic-in-L_1', 'analytic-in-L_2', ..., where L_1 and L_2 are specific linguistic frameworks. What would be needed to vindicate the notion of analyticity, so Quine apparently demands, is a definition of 'analytic' for *variable* languages L.

If this were the heart of the issue then Carnap could plausibly reject Quine's demand as unreasonable. This is in fact this what he did on one occasion:

> In case Quine's remarks are meant as a demand to be given one definition applicable to all systems, then such a demand is manifestly unreasonable; it is certainly neither fulfilled nor fulfillable for semantic and syntactic concepts, as Quine knows. (Carnap 1990: 430)

The paradigm case of a semantic concept Carnap is alluding to here is *truth* as treated by Tarski. The thought is that while Tarski's machinery allows us to characterise truth for *specific* languages, it does not give us a way to define a truth-predicate that can be applied to sentences of *arbitrary* languages. Quine doesn't seem to have a problem with Tarski's theory of truth, however, and so it is understandable that Carnap doesn't regard the fact that 'analytic' needs to be defined for each particular framework as problematic either.

That the disagreement between Carnap and Quine fundamentally concerns their different conceptions of language comes out more clearly if one looks beyond "Two Dogmas". While trying to respond to Quine's criticisms, Carnap in fact asked Quine whether the target of his objections are definitions of analyticity for natural or formal languages with explicit rules. Quine responds that for him this distinction is not essential:

> You ask whether I mean "(a) natural languages" or "(b) codified languages . . . based on explicitly formulated rules." [...] *It is indifferent to my purpose whether the notation be traditional or artificial, so long as the artificiality is not made to exceed the scope of "language" ordinarily so-called, and beg the analyticity question itself.* [...] The languages I am talking about comprise natural languages and any (used, or interpreted) artificial notations you like, e.g. that of my *Mathematical Logic* plus extra-logical predicates. They are not uninterpreted notations. (Quine, Letter to Carnap on August 9th, 1954 (Creath 1990: 438f), my emphasis)

This view, according to which natural and formal languages are continuous with each other, is a major departure from Carnap's conception. Quine already noted this crucial difference in the 1940s:

> [M]y attitude toward 'formal' languages is very different from Carnap's. Serious artificial notations, e.g. in mathematics or in your logic or mine, I consider supplementary but integral parts of natural language. [...] Thus it is that I would consider an empirical criterion [...] a solution of the problem of synonymy in general. And thus it is

> also that [...] I am unmoved by constructions by Carnap in terms of so-called 'semantical rules of a language'. (Quine, Letter to Church from August 14, 1943, quoted in Verhaegh 2018: 74f)

For Carnap the distinction between natural language, which is messy and often imprecise, and formal languages, which come with explicit rules, is very important. While it might be interesting to study natural language directly and try to come up with behavioural criteria for the application of certain concepts, this is not what Carnap is chiefly concerned with. Instead he develops formal languages in which there are clear explicit definitions for analyticity. This is not to say, however, that analyticity in formal languages is just an arbitrarily defined notion with no relation whatsoever to anything we do in natural language. Carnap rather thinks that we have a reasonably clear understanding of which natural language sentences we typically regard as analytic, even though we cannot capture this imprecise notion directly by explicitly defining it in other terms (Carnap 1990: 431). And even if this were not so, Carnap thinks that the concept of analyticity might nevertheless prove to be theoretically useful despite not corresponding to any natural language notion. He thus helpfully sums up his disagreement with Quine as follows:

> Especially Quine's criticism does not concern the formal correctness of the definitions in pure semantics; rather, he doubts whether there are any clear and fruitful corresponding pragmatical concepts which could serve as explicanda. That is the reason why he demands that these pragmatical concepts be shown to be scientifically legitimate by stating empirical, behavioristic criteria for them. If I understand him correctly, he believes that, without this pragmatical substructure, the semantical intension concepts, even if formally correct, are arbitrary and without purpose. I do not think that a semantical concept, in order to be fruitful, must necessarily possess a prior pragmatical counterpart. It is theoretically possible to demonstrate its fruitfulness through its application in the further development of language systems. (Carnap 1955: 35)

Does the Carnap-Quine debate have a winner? It seems not, since on their own terms Quine and Carnap are both vindicated. One might of course wonder whose conception of language is to be preferred, but it is not easy to see how to adjudicate such a question, since it is entangled with two very different conceptions of the role of philosophy. For Carnap one of the primary tasks of philosophy is to study and develop formal languages to be used for science and by scientists. According to Quine’s *naturalism*, on the other hand, philosophy is continuous with science, and there is thus no sense in which philosophy can stand outside of science and deal with the construction of the language of science. It is hardly clear what sort of criteria one could use to evaluate the costs and benefits of the respective conceptions.

3.2 Quine’s Triviality Objection

The summary I provided hopefully demonstrates that it is too simplistic to regard Quine as the winner of the debate over analyticity. However I think it would also be too quick to conclude that there was really no disagreement between Carnap and Quine at all, since they start from different but equally acceptable background assumptions. It is not very promising to tackle questions such as ”what should we demand of a definition of analyticity?” or ”whose conception of language is superior?” directly, since in the abstract it is difficult to state what the criteria of success are. Instead my plan to look at the one specific use Carnap has for the notion of analyticity that is the topic of this thesis: namely to declare mathematics to be true in virtue of linguistic rules. Since Quine also objected to this specific employment, it makes for a useful and more tractable case study.

As we already know, Carnap wants to define analyticity in such a way that all mathematical statements come out as either analytic, i.e. true, or contradictory, i.e. false. In light of Gödel’s incompleteness theorems this task can only be achieved by stating the definition in a metalanguage which itself contains mathematics and is stronger than the object language to whose sentences the predicate ’analytic’ will be applied. It is of course easy to feel that there is something objectionably circular about this approach. And this is indeed the kind of objection Quine raises in his ”Carnap and Logical Truth”, since he thinks that Carnap’s approach trivialises the claim that mathematics is true in virtue of linguistic rules:[3]

> So construed, however, the thesis that logico-mathematical truth is syntactically specifiable becomes uninteresting. For, what it says is that logico-mathematical truth is specifiable in a notation consisting solely of (a) [names of signs], (b) [an operator expressing concatenation of expressions], *and* the whole logico-mathematical vocabulary

[3] For further discussion of this paper see Isaacson 1992, Creath 2003 and Gregory 2003a.

> itself. But *this* thesis would hold equally if "logico-mathematical" were broadened (at *both* places in the thesis) to include physics, economics, and anything else under the sun; Tarski's routine of truth-definition would still carry through just as well. No special trait of logic and mathematics has been singled out after all. (Quine 1963: 400)[4]

This sounds like a compelling objection, but the situation is not as straightforward as it appears to be. Quine is clearly correct that Carnap's account of mathematical truth is not as interesting as the more ambitious attempt to make sense of mathematics without employing *any* mathematical concepts at all. But, as discussed in section 1.2.3, this was not Carnap's aim.

The question Quine's complaint raises is definitely a good one. For if it is fine to use a metalanguage that includes a translation of the object language, then the metalanguage for a physical theory can include physical notions. And this makes it natural to think that we can define an analyticity-predicate for physics as well, which would undermine the claim that analyticity is a *distinguishing* feature of mathematics. If so, then Quine would have identified a problem Carnap should be worried about on his own terms. For if Quine is right, then it seems that the decision to call mathematics but not physics analytic is an entirely arbitrary one, and it is hard to see why we should attach any theoretical importance to the notion of analyticity.[5]

I think, however, that without supplementation Quine fails to establish the result that Carnap's characterisation of analyticity is trivial and uninteresting. In order to see why, remember that at its core Carnap's definition of 'analytic' for mathematics amounts to the following proposal:

> Mathematical Analyticity
> ϕ is analytic **iff** $\mathbb{N} \vDash \phi$, where $\mathbb{N}$ is the standard model of arithmetic.

4 More recently, the same point has been made by Peter Koellner:

> [...] Carnap revised the account [of the distinction between logical and factual truth] by replacing this notion with the notion of a truth value being determined by semantical rules. However, this is hardly an improvement. The semantical rules are just a Tarskian truth definition which is, in effect, simply a translation of the object language into the meta-language. It is quite unclear what it means to say that these rules determine the truth value of a given statement. (Koellner ms: 34)

5 For this reason Goldfarb thinks that Carnap's acceptance of semantics was an unwise move, since the "semantical approach gives uniform treatment to the logical/mathematical vocabulary and the empirical vocabulary [... which] makes it much harder to distinguish between [...] analytic and synthetic" (Goldfarb 1997: 63).

As I understand Quine's objection, it rests on the assumption that one can equally apply Carnap's definition of analyticity to physics or any other science. But in order to do that in an analogous way, there would need to be something like a *standard model of physics*, which, like the standard model of arithmetic, determines all the truth values of physical sentences. And there does not seem to be such a thing. Of course there is what is called the standard model of particle physics, but it does not play the same role as standard models do in mathematical contexts. The standard model of particle physics describes what fundamental forces and kinds of particles exist, but the model itself, without adding any empirical hypotheses, does not settle all particular physical claims. It is a model *of* the physical world, and hence it would be odd to say that "there are 3728 neutrons" is true *in* the standard model of physics. This is a very different conception from that found in mathematics, since the standard model of arithmetic *just is* the universe of natural numbers, it does not merely describe it.

This point can be put in another way. A metalanguage for mathematics will include mathematical concepts, and one for physics will contain physical concepts – so far so analogous. But as Carnap sets things up, the mathematical metalanguage will be such that it settles all the truth *values* of the mathematical sentences, whereas – at least for some cases – the physical metalanguage merely states truth-*conditions* in a disquotational way:

> '$\Psi(x, y, z, t)$' is true if and only if space-time point $< x, y, z >$ instantiantes the fundamental property Ψ at t.

In this case we will hence not be able to settle whether some specific claim such as '$\Psi(a, b, c, d)$' is true or false merely by drawing on the rules stated in the metalanguage.

In a dry but important passage from his *Introduction to Semantics*, Carnap spells out what the claim that some sentence is analytic (here called L-true) in a framework (here called semantical system) amounts to in a precise way:

> In our previous discussion we found a characteristic feature of the L-true sentences of a semantical system *S*, their truth follows from the rules of *S* alone. This characterization as it stands cannot be taken as a definition of 'L-true in *S*'. If we expand the phrase 'the truth of $\mathfrak{S}_i$ follows from the semantical rules of *S*', we see that it does not belong to the metalanguage *M*, in which the definition 'L-true in *S*' has to be formulated, but to the metametalanguage *MM*, i.e. the language in which the rules for *M* are formulated. [...] we may reformulate the

> above phrase in this way: "The sentence '$\mathfrak{S}_{\iota}$ is true in S' is L-true in M". This phrase, however, speaks about M and hence belongs to MM but not to M. Therefore it cannot be taken as a definiens for '$\mathfrak{S}_{\iota}$ is L-true in S'. It rather expresses a requirement which must be fulfilled for all sentences of S if the term 'L-true in S' is to be in agreement with our intention and traditional use, or, as we may say briefly, if the definition of 'L-true in S' is to be accepted as adequate. Therefore we shall formulate the requirement as a definition (in MM) of adequacy (in M). (Carnap 1942: 83f)

This can be summed up as follows: a sentence S of some object language is analytic if and only if the truth of S can be established in the metalanguage using *only* the linguistic rules of this metalanguage, i.e. without relying on any empirical facts. And this conception precisely matches the distinction between mathematics and physics I described above. As we saw in section 2.1.2, Carnap considers the specification of the standard model of arithmetic to be a linguistic rule, and therefore the truth values of mathematical sentences follow from the linguistic rules of the metalanguage – hence they count as analytic. For physics, on the other hand, the metalanguage merely gives disquotational truth-conditions *using* physical vocabulary, and hence the truth values of the object language statements will in many cases depend on empirical factors. Physics thus counts as synthetic.

I therefore think that Quine overstates his case in the earlier quote. The considerations he gives do not establish that Carnap's distinction is completely arbitrary and therefore trivial, and so Quine's advice to "stop tugging at our bootstraps altogether" (Quine 1951: 33) seems premature. In fact, Quine himself seems to have had some reservations about the effectiveness of his arguments. "Carnap and Logical Truth" begins with a striking caveat:

> My dissent from Carnap's philosophy of logical truth is hard to state and argue in Carnap's terms. This circumstance perhaps counts in favor of Carnap's position. (Quine 1963: 385)[6]

It is therefore not surprising that in his reply from the Schilpp-volume, Carnap struggles to see which aspect of his position Quine wants to reject for what reason. In an uncharacteristically sarcastic passage, Carnap writes the following:

6 Quine's paper was originally written in 1954, but since the publication of the Schilpp-volume was delayed by many years Quine in the meantime published it in *Synthese* (Quine 1960a). This version, as well as later reprints, do not contain the preamble I am quoting.

> [... Quine] himself says soon afterwards: ”I do not suggest that the linguistic doctrine is false”. I presume that he wants to say that the doctrine is not false. (If so, I wish he had said so!) He nowhere says that the doctrine is meaningless; [...] Therefore we may presume that he regards the doctrine as true. (If so, ...!) The main point of his criticism seems rather to be that the doctrine is ”empty” and ”without experimental meaning”. With this remark I would certainly agree, and I am surprised that Quine deems it necessary to support this view by detailed arguments. (Carnap 1963b: 917)

It is easy to sympathise with Carnap's reaction here. For while Quine's arguments may work against *some* ways of drawing the analytic/synthetic distinction, they do not seem very effective against the way *Carnap* makes use of it, and consequently it becomes hard to see the relevance of many of Quine's points. We thus end up in a similar situation as we did in the previous chapter: just as Gödel's argument based on the incompleteness theorems was shown to rely on un-Carnapian background assumptions, the same seems to hold for Quine. There remains the nagging feeling that Gödel and Quine have correctly identified *some* deep problem for Carnap's position, but we have not yet found a clear formulation of it.

In the next part of the thesis I will try to improve on this messy situation by showing that there *is* a way to argue against Carnap's claim that mathematics is analytic without begging any questions. The argument I ascribe to Beth combines elements of both Gödel and Quine. Like the argument from consistency it is based on the incompleteness theorems, and like Quine it makes use of the idea that, on cost of triviality, there need to be behavioural criteria for the use of 'analytic'. In the next chapter I will start by giving my own reconstruction of Beth's argument, before defending it against alternative readings in the two subsequent chapters.

Part II

Beth versus Carnap

Chapter 4

Beth's Argument from Non-Standard Models

In this chapter I introduce the hero of my thesis: the Dutch philosopher and logician E. W. Beth. My aim is to show that what I call his *argument from non-standard models* successfully demonstrates that Carnap's response to Gödel's incompleteness theorem fails. Rightly understood, so I will claim, Beth's argument is more forceful than both Gödel's own argument from incompleteness and Quine's influential objections to the notion of analyticity.

Beth's own presentation of his argument is not straightforward, and so the first half of the chapter contains a detailed reconstruction of Beth's line of reasoning, as well as discussion of Carnap's response to it. In the course of this we will revisit an issue that already arose in previous chapters, namely the relationship between linguistic frameworks and natural language. In the second half of the chapter I then argue that Carnap's response to Beth is unsatisfactory. I demonstrate this by focusing on two of the central tenets of Carnap's philosophy: the principle of tolerance and the internalist treatment of abstract objects. In the end we will see that Beth's controversial claim that Carnap cannot treat all ontological questions as merely pragmatic can be vindicated.

4.1 Carnap and Carnap*

4.1.1 Introducing Carnap*

The argument from non-standard models is contained in Beth's contribution to the Schilpp-volume on Carnap, whose unexciting title is "Carnap's Views on the Advantages of Constructed Systems over Natural Languages in the Philosophy of Science" (Beth 1963). It is a relatively detailed study of a number of aspects of Carnap's philosophy, and also covers topics like the development of Carnap's views on mathematics from his days as a student of Frege to the writing of *Logical Syntax*, and the consequences of his acceptance of semantics.

For our purposes the relevant sections are VI, VII, and the usually ignored section X on "Ontological Commitments". Broadly speaking, Beth suggests that the possibility of interpreting formal languages in a model-theoretically non-standard way poses a problem for Carnap's philosophy of logic and mathematics. More specifically, Beth claims that this possibility shows the following:

> Beth's Thesis
> Carnap needs to rely on an intuitive interpretation of the metalanguage in *The Logical Syntax of Language* that is not fully captured by the explicitly formulated rules. This entails that the Principle of Tolerance cannot be maintained without restrictions, and also that not all ontological questions can be treated as matters of pragmatic decision. (after Beth 1963: 479f, 500f)

It would obviously a major blow to Carnap if Beth's Thesis were correct, for both the principle of tolerance and the rejection of ontological questions are core tenets of his position.

In this chapter I will use *first-order* Peano arithmetic (PA) as the example theory to reconstruct Beth's argument, since it is much more familiar than Carnap's language II towards which the original paper is targeted. The stress on first-order is important, for one may well think that many of the problems Beth diagnoses can be avoided by moving to second-order Peano arithmetic (PA2). I think that this is ultimately incorrect, but postpone further discussion to section 7.3. In order not to get lost it in the details of Beth's presentation it makes sense to first describe the overall strategy in broad strokes. The main result Beth relies on is that theories can be interpreted in unintended ways. He introduces the following distinction:

> In a treatise such as *Logical Syntax*, natural language can be used in two different ways, which I should like to denote as *strict usage* and *amplified usage*, respectively. In strict usage of natural language, we refer to a definite model of the theory to which our statements belong; it is this model which has been called the *intuitive model*. In amplified usage of natural language – and in *all* usage of formalised languages

> – on the other hand, we refer to *any* model of this theory. (Beth 1963: 479f)

Let us apply this to Peano arithmetic. Users of PA will usually want to use this system in a strict way, i.e. with a particular fixed interpretation in mind, namely the *standard model of arithmetic*.[1] But it is a consequence of both the Löwenheim-Skolem theorem and Gödel's incompleteness theorems that PA also has models which are usually called non-standard since they are not isomorphic to the intended interpretation. As we will soon see, one could in principle imagine someone using PA in a strict way with a fixed but *non-standard* model in mind. And finally there's amplified usage, in which theories are treated purely algebraically, without a particular intended interpretation, as when Hilbert writes to Frege that the axioms of geometry might apply to point and lines, but also to "love, law, chimney sweep" (Hilbert, letter to Frege from 21.12.1899, Frege 1980: 40).[2]

The important observation is that often we *do* want to use a formal system in a strict way, but that the intended interpretation is not *determined* purely by the axioms and inference rules. Based on this Beth argues as follows:

(1) What Carnap says in *Logical Syntax* must be read as involving strict usage of language, i.e. Carnap has one particular intended interpretation in mind.

(2) This intended interpretation is not pinned down by the explicit statements Carnap makes in *Logical Syntax*: someone could read the book but interpret it in an unintended way.

(3) This shows that there is a sense in which Carnap cannot replace all appeals to an intuitive notion of interpretation by explicit rules.

In order to make these points Beth introduces the fictitious protagonist Carnap*, who reads Carnap's *Logical Syntax* but interprets it in an unintended way. (Beth's

---

[1] The notion of a standard model of arithmetic is potentially ambiguous: one might mean precisely the natural numbers, in which case there is *exactly one* standard model, or count any model that is *isomorphic* to the natural numbers as standard. As I discuss in section 5.1.1, Carnap's approach only requires pinning down a model which is standard in the second, more liberal sense. Throughout the text I also assume that arithmetical terms and functions are interpreted in the usual way.

[2] In *Word and Object* (Quine 1960b: 273) Quine complains about a passage by Russell in which the latter defines mathematics as the subject in which "we never know what we are talking about, nor whether what we are saying is true" (Russell 1918: 75). At this point Russell presumably thought that it is characteristic of mathematics that it is only used in what Beth calls an amplified way.

discussion is targeted towards Carnap's language II, but I will continue to use PA.) Beth begins by making the uncontroversial point that PA has non-standard models in which some sentences get assigned different truth value from the ones they have in the standard model. Calling the standard model of PA M, he generates such a non-standard model M^* by adding $\neg \mathrm{Con}_{PA}$ to the axioms of PA. He then describes Carnap* as a logician "whose logical and mathematical intuitions are in accordance with model M^*" (Beth 1963: 478). The main task of this section will be to understand what exactly Beth means by this.

One is tempted to construe the scenario as follows: Carnap and Carnap* both look at the axioms of PA. Carnap interprets those as being about the natural numbers, while Carnap* thinks that they should be interpreted with respect to the non-standard model M^*. If they consider the truth of Con_{PA}, for instance, they will have different opinions. This simple interpretation, however, is at odds with some of the subsequent remarks Beth makes. Consider the following quote, where W_{II} is the consistency sentence for language II:

> We made an error in supposing that, for Carnap*, all theorems of II* are intuitively true; for Carnap*'s set of all theorems of II* does not coincide with our set of all theorems of II*. Roughly speaking, Carnap*'s set contains *more* theorems than ours, and for some of these additional theorems M* is not a model.
>
> It is easy to see, that for Carnap* *non*-W_{II} is a theorem of language II; hence, for him, II* coincides with II and S* coincides with S. (Beth 1963: 478)

Language II here corresponds to PA, while language II* is PA+$\neg \mathrm{Con}_{PA}$. Beth therefore says that Carnap and Carnap* *disagree* about what the theorems of PA+$\neg \mathrm{Con}_{PA}$ are, and, more specifically, Carnap* thinks that PA+$\neg \mathrm{Con}_{PA}$ has *more* theorems than Carnap supposes. On the face of it these claims are hard to make sense of. For while initially Carnap and Carnap* were introduced as differing on the *semantic* interpretation of certain formal theories, they are now presented as differing on merely *syntactic* properties as well.

Beth must therefore have something more than the straightforward interpretation I gave in mind, as Carnap himself helpfully clarifies in his own response to Beth:

> But now Beth proceeds to make a number of further statements about Carnap* which at first glance appear as obviously false, e.g., the statement that the set of all axioms of II* is different for Carnap* and for

> us, i.e. Carnap and Beth, and the statement that for Carnap* the languages II* and II coincide. These statements certainly do not follow from the sole assumption that Carnap* applies the interpretation *Int** to the object languages. (Carnap 1963b: 928f)

Carnap also sums up the solution to these initially puzzling remarks in a more concise way than Beth himself:

> Beth's statements are understandable only on the basis of an additional assumption, namely that Carnap* interprets not only the symbolic object languages but also the metalanguage *ML* in a way different from Carnap. Therefore I suppose that Beth makes this additional assumption, although he does not state it explicitly. And I presume, more specifically, that Beth assumes that the interpretation of *ML* by Carnap* is analogous to his interpretation *Int** of II (and II*). (Carnap 1963b: 928f)

This seems correct: for Beth's description to make any sense, it must be the case that the metalanguage Carnap* uses is non-standard in some way. In particular, Carnap*'s metalanguage *ML** must be such that in it Carnap* can conclude that PA is inconsistent. For then it would be clear why he cannot distinguish between PA and PA+$\neg Con_{PA}$: adding a sentence to an inconsistent theory just gives us the same inconsistent theory once again.

The question then becomes how exactly Carnap* manages to prove the inconsistency of PA in his metalanguage though. As Beth describes it, when Carnap and Carnap* both go through *Logical Syntax*, they will agree up to the following point:

> The first place where real trouble arises is indicated by Carnap himself on page 113. Carnap there points out that a certain point in the given definition of 'analytic in II' may appear dubious. This definition contains certain phrases meaning "for all syntactical properties of accented expressions ..." Now the meaning of such phrases for Carnap and for Carnap* is different, as *for Carnap* the set of all accented expressions is larger than it is for Carnap*. It follows that Carnap will make (or accept) certain statements concerning the set of all syntactical properties of accented expressions which Carnap* must reject. (Beth 1963: 480f, my emphasis)

Beth then writes that this difference in how to interpret the meaning of 'for all syntactical properties of accented expressions' has the consequence that Carnap* will reject Theorem 36.6 of *Logical Syntax*. And this theorem is the statement that Con_{PA} is true even though not derivable, which we already discussed in section 2.1.1 when explaining how Carnap intends to overcome Gödelian incompleteness.

Why does Carnap* disagree with Carnap on this point? Remember that the consistency sentence Con_{PA} is the following universal claim:

$$\forall x \neg Pr_{PA}(x, \ulcorner \bot \urcorner)$$

And although its instances '$\neg Pr_{PA}(0, \ulcorner \bot \urcorner)$', '$\neg Pr_{PA}(0', \ulcorner \bot \urcorner)$, ... are derivable in PA, Con_{PA} is not. Carnap solved this problem by having a rule in the metalanguage that allowed one to conclude that a universal generalisation is analytic, and hence true, if all its instances are analytic. Earlier we already noted that in giving this rule Carnap relied on an infinite class of accented expressions or numerals 0, $0'$, $0''$, ..., which he assumes to be isomorphic to the standard numbers.

It is precisely this last point that Carnap* takes issue with. For he interprets the range of the quantifiers differently from Carnap, and admits some numerals as substitution instances that are non-standard in that they are not generated in a finite number of steps starting from '0'. And not only does Carnap* believe that *there are* such non-standard numerals, but he furthermore holds that one of them encodes the proof of a contradiction from the axioms of PA. In his metalanguage *ML**, Carnap* is thus able to show that for some non-standard numeral ξ, $Pr_{PA}(\xi, \ulcorner \bot \urcorner)$ is analytic, and hence he holds the universal claim Con_{PA} to be *false*.

The case of Carnap* will appear very peculiar, for it seems that his belief that there is an instance '$Pr_{PA}(\xi, \ulcorner \bot \urcorner)$' does not follow from any other general principles he accepts, but is rather an independent and brute conviction. Beth happily admits that this is a strange – even 'psychopathic' (Beth 1963: 484) – opinion to have, in particular since Carnap* is not able to actually produce a syntactic proof of a contradiction from the axioms of PA (Beth 1963: 481). But, importantly, while it is psychologically implausible that someone like Carnap* actually exists, the scenario is not *incoherent*.

There are admittedly some other passages in Beth's paper that cast doubt on whether the scenario he envisaged is precisely the one I have described here, and I will come back to these in section 5.1.3. But for now the more important question is why the case of Carnap* should amount to an *objection* to Carnap's position at all.

4.1.2 What's the Problem?

Carnap* has now been introduced in great detail. But what consequences can be derived from the possibility of such a character? According to Beth the implications are substantial indeed, for he takes himself to have established the following three theses:

(B_1) Carnap needs to rely on an intuitive interpretation of the metalanguage in *The Logical Syntax of Language* that is not fully captured by explicit rules (Beth 1963: 479f).

(B_2) The Principle of Tolerance cannot be maintained without restrictions (Beth 1963: 479).

(B_3) Carnap cannot treat all questions concerning ontology as a merely pragmatic matter (Beth 1963: 500f).

Why (B_1) follows from the possibility of Carnap* is relatively easy to see. If one thought that the explicitly stated rules that of *Logical Syntax* uniquely determine a particular intended interpretation, then Beth successfully shows that this is not the case. This part of Beth's argument has also been discussed in the secondary literature, and in principle commentators tend to agree with Beth's conclusion:

> Clearly, if the metalanguage is a rich one, and if our understanding of it cannot be exhaustively explicated in terms of rules, deductive procedures in axiomatic systems, or the like, then Carnap's "presupposition" is an admission that much can never be made explicit, but must simply be tacitly relied upon. This fits poorly with Carnap's proclaimed standards of exactitude and rigor. (Goldfarb and Ricketts 1992: 72)

> [...] since no amount of purely syntactical or inferential behavior will guarantee that [... two investigators] share a metalanguage, they must trust their practical identifications of shared vocabulary and inference rules. This "mystical" trust seems in tension with Carnap's recommendation that we construct language systems whose logical syntax is fixed and unambiguous. (Ebbs 2017b: 31)

What becomes apparent in these representative passages, however, is that Beth's objection has not been regarded as a deep challenge that goes at the heart of Carnap's view. Instead, Beth has largely been read as flagging that Carnap overstates his case when he exclusively praises the virtues of exact rules. This is of course not ideal, but more a case of misleading advertising than a challenge to the coherence of his position.

For Beth's argument to have more bite there needs to be a way to get to claims (B_2) and (B_3) based on (B_1), for these are more damaging to Carnap's position. Unfortunately Beth's own arguments relating to (B_2) and (B_3) are very brief and elusive, and so the connection between the need for an intuitive interpretation, and the principle of tolerance and ontological questions, has been largely ignored. In the second half of this chapter I will show that it is possible to defend these further claims, but some build-up is required before we get there, since the dialectical situation is complex. Unlike with Gödel's "Is Mathematics Syntax of Language?" we are in the lucky position to have a reply to Beth's arguments from Carnap himself, and it will be helpful to discuss this first.

Carnap begins by agreeing with one of Beth's main points, namely that when using a language as a metalanguage, we need to do so with a particular intended interpretation:

> [...] Beth's thesis says that it is essential for the purposes of my theory that the English words of my metalanguage *ML* are sometimes used with a fixed interpretation. I emphatically agree; I would even say that this is the case not only sometimes but practically always. For the reasons explained earlier, this seems to me so obvious that I am surprised that Beth should regard it as necessary to demonstrate it by particular examples. (Carnap 1963b: 930)

Furthermore, Carnap is also happy to grant that someone like Carnap* who uses an unintended metalanguage *ML** is *possible*. He does not regard this possibility as a reason to worry, however, for he thinks that it is reasonable to assume that he and the readers of his book use a metalanguage with the standard interpretation of syntax, i.e. something like *ML*. The assumption, in other words, is that *we* are like Carnap, whereas Carnap* is just a fiction:

> Since the metalanguage *ML* serves as a means of communication [...] I always presupposed both in syntax and in semantics that a fixed interpretation of *ML*, which is shared by all participants, is given. This interpretation is usually not formulated explicitly; but since *ML*

> uses English words, it is assumed that these words are understood in their ordinary senses. The necessity of this presupposition of a common interpreted metalanguage seems to me obvious. (Carnap 1963b: 929)

This of course raises the question of how we managed to arrive at the intended interpretation rather than a devious one in the first place. Carnap admits that there is no fool-proof method, but points out that it would be overdemanding to require that any misinterpretation of *Logical Syntax* is *impossible*. It is for instance very important that readers of *Logical Syntax* do not interpret the expression "no occurrence" as meaning "at least one occurrence" (Carnap 1963b: 929), but surely the mere possibility that someone *could* read the book in such an unintended way does not amount to an objection.[3] What matters is that we do not *actually* misinterpret each other in the way Beth envisages, and the possibility of misinterpretation in itself cannot undermine *our* shared understanding of the metalanguage *ML*.

Let us call this response the *we actually speak ML* reply. On the face of it seems extremely simple, and one might worry that Carnap fails to fully engage with Beth's reasoning. As I will argue later this is true to some extent, but it is important to appreciate that the reply is a very powerful move. In fact, I think that *if* Carnap were justified to maintain that we actually speak Carnap's *ML* rather than Carnap*'s *ML**, he would also be in a position to set Beth's concerns about the principle of tolerance and ontology aside. This also demonstrates, however, that to presuppose that we speak metalanguage *ML* rather than *ML** is not as innocent as Carnap makes it out to be. He writes as if this fact is as uncontentious as the fact that we all speak some natural language such as English or German, which is a misleading analogoy. In the end I will maintain that Carnap is *not* in a position to hold on to the crucial presupposition of his reply.

That Carnap doesn't mention the principle of tolerance in his reply anywhere is a bit surprising, given that it is so central to his philosophical outlook. One can try to reconstruct why he did not take Beth's claim that the principle must be

[3] Many years later Paul Benacerraf made a similar point with the help of T. S. Eliot:

> ...I gotta use words when I talk to you
> But if you understand or if you don't
> That's nothing to me and nothing to you
> We gotta do what we gotta do ...
> (T.S. Eliot, quoted in Benacerraf 1985: 111)

restricted seriously, however. One of Beth's most explicit passages in this regard is the following:

> It should be noted that we also meet here with a limitation regarding the Principle of Tolerance. Indeed, Carnap could be tolerant with respect to Carnap*, for Carnap would be able to understand why Carnap adopting for certain personal reasons additional axioms for Language II is compelled to accept additional theorems and to reject certain (and indeed all) models for language II. But Carnap* would never be able to understand why Carnap, having accepted certain axioms and certain rules of inference, as stated in Logical Syntax, stubbornly refuses to accept [the negation of the consistency sentence] non-W_{II} as a theorem and believes Language II to have a model. (Beth 1963: 478)

This elucidates what the connection to the possibility of non-standard interpretations is: namely that the principle of tolerance *itself* can be interpreted non-standardly, since it relies on a particular conception of what syntax is. Carnap demands that logicians "give syntactical rules instead of philosophical arguments", but the case of Carnap* shows that someone's understanding of syntax might diverge from ours.

It is tempting to construe Beth as making the following point: if we want to follow the principle of tolerance we need to make sure that we share a common understanding of syntax, for otherwise inexplicable disagreements like that between Carnap and Carnap* will arise. And if this is the only upshot of Beth's paper, then Carnap can just agree with its conclusion. It is clear from his reply that he thinks that we *do* share a common conception of syntax, and hence encounters like that between Carnap and Carnap* do not *actually* arise. Carnap will admit that they are possible, but it seems fair to say that such cases do not really demonstrate a *limitation* of the principle of tolerance in any problematic sense. For it would surely be unreasonable to demand that there is no possibility of misinterpreting the principle, and have it resolve *any* case of miscommunication. Carnap thus has good reasons to set aside Beth's (B_2) as non-threatening.

This response relies on the *we actually speak ML* reply, however, and thus demonstrates how important this assumption is for Carnap's overall attitude towards Beth's arguments. In the next section I will scrutinise what, from Carnap's perspective, it even means to claim that we speak a particular metalanguage. This will then enable me to defend Beth's objections by arguing against the *we actually speak ML* reply.

4.1.3 Metalanguages and How to Use Them

We can approach the topic of using metalanguages by looking into Beth's claim that the principle of tolerance faces certain limitations some more. In general Beth is sympathetic towards Carnap's tolerant attitude concerning different systems of logic, so his objection to the principle is not that Carnap fails to recognise that there is one true logic. After expressing this partial agreement with Carnap, Beth continues as follows:

> But there are, in my opinion, certain natural limitations which go farther than those which Carnap would be ready to accept. And this is in particular the case, if the language under consideration is to be used as a metalanguage. (Beth 1963: 498)

I think that this innocent remark is of great importance, for reflecting on why the use of a certain language *as a metalanguage* should be more problematic than other uses will give us the clue to properly understanding the challenge posed by Carnap*.

Carnap's response to Beth was that we do not need to be worried about Carnap*, since the metalanguage we actually speak is *ML* and not *ML**. This sounds straightforward at first, until we remember some of *ML*'s features. In *ML* we were able to establish that Con_{PA} is true using a non-recursive and infinitary rule, for instance. Does it make sense to say that the language we actually speak – which is first and foremost a natural language such as English or German – has this ability as well? This question is hard to answer, since natural languages seem quite different from formal languages with explicit rules such as *ML*. Carnap's reply thus needs to be reconsidered. First we need to understand what it even means to say that we *use* a formal system like *ML* as the metalanguage. And, secondly, we then need to assess whether Carnap's claim that we in fact do so is plausible in light of the answer to the first question.

In order to proceed we need to look at the relationship between natural language and linguistic frameworks in more detail. In most of his work Carnap focusses exlcusively on the formalised frameworks with explicit rules, but his "Foundations of Logic and Mathematics" contains a discussion of how one can *use* such frameworks. It becomes clear that Carnap understands natural languages in a broadly behaviouristic way:

> A language, as, e.g., English, is a system of activities or, rather, of habits, i.e., dispositions to certain activities, serving mainly for the

> purposes of communication and of co-ordination of activities among the members of a group. (Carnap 1939: 3)

Obviously natural language understood in this sense cannot be the same kind of thing as linguistic frameworks, since the latter are constituted by linguistic rules. But the two notions of language are not completely unrelated either. Carnap thinks that we can *coordinate* languages understood as formal calculi with languages understood as systems of dispositions. He describes this process of coordination as a kind of radical translation, where we observe a population of speakers and their behaviour, and then construct a linguistic framework with rules that correspond to the speakers' linguistic dispositions.[4] He stresses that the interpreters should not think of themselves as literally investigating the rules of the actually spoken natural language, but rather as offering a rational reconstruction:

> Strictly speaking, the rules which we shall lay down are not rules of the factually given language B; they rather constitute a language system corresponding to B which we will call the semantical system B-S. (Carnap 1939: 6f)

Despite this caveat we can now give some content to the claim that we, or some other group of people, speak a particular formal language or adopt a linguistic framework. For it amounts to the claim that the explicit linguistic rules of the relevant framework correspond to the linguistic dispositions the speakers actually have. If, to take a simple example, speakers are disposed to reason in accord with modus ponens, then it is appropriate to capture this fact using a calculus that contains modus ponens as an inference rule.

The situation is complicated by the fact that our linguistic behaviour does not unambiguously determine one unique framework to be the best reconstruction. As Carnap points out, the dispositions of speakers can usually be captured by various competing framework rules:

> Suppose we have found that the word 'mond' of B was used in 98 per cent of the cases for the moon and in 2 per cent for a certain lantern. Now it is a matter of our decision whether we construct the rules

4 I rely on Ricketts' view that for Carnap logical calculi need to be coordinated with the 'logically amorphous' natural language (Ricketts 2003). This interpretation has been challenged by Carus, who holds that instead of a sharp divide between natural and formal languages Carnap sees their relationship as *continuous* (Carus 2004, Carus 2007, for discussion see Richardson 2012 and Wagner 2012). I discuss Carus's views some more in section 5.2.3.

> in such a way that both the moon and the lantern are designata of 'mond' or only the moon. If we choose the first, the use of 'mond' in those 2 per cent of cases was right – with respect to our rules; if we choose the second, it was wrong. (Carnap 1939: 6)

It is tempting to conclude that this underdetermination of frameworks by linguistic behaviour leads to the kind of problems Kripke's interpretation of Wittgenstein's rule-following considerations brings out (Kripke 1982). Suppose we observe a population of speakers who seem to use addition, for instance, but never deal with numbers greater than 10,000. It might seem that ascribing a framework that includes rules for addition to them would be as correct as the ascription of a deviant framework with rules for *quaddition* based on the following quus-function:

$$x \text{ quus } y = \begin{cases} x + y & \text{if } x \leq 10{,}000 \text{ and } y \leq 10{,}000 \\ 137 & \text{otherwise} \end{cases}$$

What would Carnap say about this case? At times he seems to endorse the radical thesis that there are *no* standards of correctness at all when it comes to coordinating linguistic frameworks with speech behaviour:

> The facts [about linguistic behaviour] do not determine whether the use of a certain expression is right or wrong but only how often it occurs and how often it leads to the effect intended, and the like. A question of right or wrong must always refer to a system of rules. (Carnap 1939: 6f)

In a way this would fit Carnap's attitude to framework choice in general: just as there are no non-pragmatic standards of correctness when it comes to framework adoption, there are no factual external questions about framework coordination either. But this view has radical consequences, for we seem forced to say that there are *no* semantic facts about natural language at all – not even facts like *in German, the word 'Mond' refers to the moon*.

In Carnap's "Meaning and Synonymy in Natural Languages" from 1954, it becomes clear that this radical non-factualism is not what Carnap has in mind.[5] Carnap there argues that while the extension of some terms may not be completely precise, there are nevertheless empirical facts about what natural lan-

[5] Some commentators suggest that Carnap *should* have embraced this option, however, and I will hence come back to this topic in in chapter 6.

guage expressions refer to (Carnap 1955: 34). More than that, he even intends to defend the existence of facts about *intensions* against the criticisms of Quine:

> The *intensionalist* thesis in pragmatics, which I am defending, says that the assignment of an intension is an empirical hypothesis which, like any other hypothesis in linguistics, can be tested by observations of language behavior. On the other hand, the *extensionalist thesis* asserts that the assignment of an intension, on the basis of the previously determined extension is not a question of fact but merely a matter of choice. The thesis holds that the linguist is free to choose any of those properties which fit to the given extension; he may be guided in his choice by a consideration of simplicity, but there is no question of right or wrong. Quine seems to maintain this thesis [...] (Carnap 1955: 37)[6]

In general Carnap thus seems happy to accept semantic facts about natural language, which is good since they are required by his reply to Beth – that we speak metalanguage *ML* rather than *ML** – as well.

For the purposes of my argument I will also assume that Carnap can respond to Kripke's challenge somehow, and maintain that our linguistic behaviour is best captured by an addition rather than a quaddition framework, by a modus ponens framework rather than some deviant alternative inference rule, and so on. Whether this assumption is justified in light of Carnap's actual view is admittedly doubtful. If there is any difference between a population that uses addition and one that uses quaddition even though they never actually calculate with numbers larger than 10,000, this difference must presumably consist in the presence of certain *dispositions* to calculate in such-and-such ways. And Carnap's theory of dispositions makes it quite hard to account for dispositions of this kind. In "Testability and Meaning" he explains disposition predicates such as 'x is soluble' in terms of so-called *reduction sentences*:

> 'Q_3' cannot be defined by D, nor by any other definition. But we can introduce it by the following sentence:
>
> (R:) $(x)(t)[Q_1(x, t) \supset (Q_3(x) \equiv Q_2(x, t))]$,

6 On another occasion, Carnap also writes that "Quine's arguments to the effect that the lexicographers actually have no criterion for their determinations [of the meanings of words] did not seem at all convincing to me" (Carnap 1963b: 920).

> in words: "if any thing x is put into water at any time t, then, if x is soluble in water, x dissolves at the time t, and if x is not soluble in water, it does not." This sentence belongs to that kind of sentences which we shall call reduction sentences. (Carnap 1936: 440f)

Reduction sentences are in effect conditionals which, even though they cannot be used to *eliminate* the disposition predicate as an explicit definition would, state in what circumstances the disposition predicate does or does not apply. The reduction sentences usually involve a *test condition*, in this case 'x is put in water'. And as Carnap himself stresses, one consequence of his account is that for many objects it will be *indeterminate* whether they are soluble or not, since they are never put into water at all:

> If we establish one reduction pair (or one bilateral reduction sentence) as valid in order to introduce a predicate 'Q_3', the meaning of 'Q_3' is not established completely, but only for the cases in which the test condition is fulfilled. [...] In the case of the predicate 'soluble in water' we may perhaps add the law stating that two bodies of the same substance are either both soluble or both not soluble. This law would help in the instance of the match; it would, in accordance with common usage, lead to the result "the match c is not soluble," because other pieces of wood are found to be insoluble on the basis of the first reduction sentence. Nevertheless, a region of indeterminateness remains, though a smaller one. (Carnap 1936: 445)

By the same kind of reasoning, so it seems, it would be indeterminate whether 'x is disposed to respond 137 when asked to calculate with number larger than 10,000' applies to anyone who has never actually done so – at least unless we can find any general laws of nature which settle this matter.[7]

We will revisit Carnap's account of dispositions in section 7.4, but for now I will make the dialectically generous assumption that he can respond to rule-following skepticism à la Kripkenstein in some way. The point I will make in the following is that the case of metalanguages *ML* versus *ML** is relevantly unlike the case of addition versus quaddition. The assumption that Carnap has a solution to the rule-following problem on its own thus does not enable him to rebut the challenge posed by Beth.

7 In Carnap 1955 Carnap sounds very optimistic about invoking general laws of nature to account for dispositions about linguistic behaviour. For a more wide-ranging survey of Carnap's position on dispositions see Schmitz 2007.

4.2 Tolerance and Abstract Objects

4.2.1 Against Infinitary Rules

I am now in a position to present my interpretation of Beth's argument. In this section I will give a reading of Beth's second thesis that turns it into a real challenge to Carnap:

> (B_2) The Principle of Tolerance cannot be maintained without restrictions (Beth 1963: 479).

The line of thought can be summed up as follows: even if we grant that linguistic dispositions can be coordinated with framework rules for cases like modus ponens or addition, this strategy does not suffice to substantiate the claim that we speak *ML* rather than *ML**. This is because these metalanguages contain non-recursive inference rules, and, given some natural assumptions, there are no linguistic dispositions that these rules could correspond to. When it comes to metalanguages, tolerance thus needs to be restricted to languages with *recursive* rules – which undermines Carnap's approach to the philosophy of mathematics.

Remember that the aim of the strong metalanguage in *Logical Syntax* is to define *analyticity* in such a way that each purely mathematical sentence is either analytic or contradictory. And it is clear that there is no straightforward sense in which we have dispositions that correspond to this feature of the metalanguage, as Ricketts also points out:

> It is implausible to hold that a classical mathematician who uses Carnap's Language II [...] is in any sense disposed to affirm or deny each mathematical sentence of this language. And nothing Carnap says suggests he believes mathematicians are so disposed. Speech habits thus do not fix the L-rules for a calculus instantiated by a language. (Ricketts 2003: 262)

One can argue, moreover, that this lack of a disposition to assert all true mathematical sentences is not just a contingent fact, but that – for us finite human beings at least – it is *impossible* to have such dispositions. While it is relatively uncontroversial to think that there are dispositions corresponding to finitary and recursive inference rules such as modus ponens, it is widely held that human beings do *not* have the capacity to follow non-recursive and infinitary rules such as the ω-rule. This has been called the Cognitive Constraint: "humans cannot be attributed non-computational causal powers" (Warren and Waxman 2020: 485). It is motivated by McGee in the following way:

> Human beings are products of nature. They 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 simulable by a Turing machine. This will hold even for idealized humans who never make mistakes and who are allowed unlimited time, patience, and memory. (McGee 1991: 117)

To ascribe the ability to assert all true mathematical sentences to someone would violate the Cognitive Constraint. For since no Turing machine can enumerate all the true mathematical sentences, it would be impossible to model the relevant ability. And this observation also casts doubt on a claim by Carnap we already encountered in section 2.1.2, namely that "there is nothing to prevent the practical application of" infinitary rules such as the ω-rule. Whatever Carnap had in mind when he wrote this, there is a clear sense in which we *cannot* just use the ω-rule like any other rule. For Turing machines cannot use rules with infinitely many premises, and if the Cognitive Constraint is correct, then neither can we.[8]

The natural conclusion is that our dispositions can at most correspond to recursive inference rules of linguistic frameworks, not to non-recursive ones. But since the latter are crucial to establish the analyticity of mathematics in the metalanguage *ML*, this undermines the claim that we actually speak this particular metalanguage.

Ultimately I think that this conclusion is not only natural, but indeed correct. When read in the way I propose, Beth's argument successfully establishes that Carnap's reliance on non-recursive inference rules is in tension with his own conception of language. This claim can be further substantiated by considering Ricketts' reaction to the lack of dispositions corresponding to non-recursive rules. He suggests that there is nevertheless a sense in which our speech behaviour can match a non-recursive frameworks, even though the relevant kind of agreement is quite loose:

> We see, then, that as regards transformation rules, the agreement between a calculus and speech habits in virtue of which a language

[8] This is at least the orthodox view among philosophers of mathematics (Field 1994, Raatikainen 2005, Smith 2013: 332, Button and Walsh 2018: chapter 7). In section 7.4 I discuss Jared Warren's recent argument that following the ω-rule is possible without violating the Cognitive Constraint.

> can be taken to instantiate a calculus is rather loose. For a speaker's habits to agree with a calculus, Carnap appears to require little more than that the speaker not be disposed to affirm any contravalid [= false] sentence nor to deny any valid [= true] sentence. (Ricketts 2003: 262)

Is it possible to argue that our speech dispositions correspond to *ML* rather than *ML** on this basis? At first sight one might think that the answer is *yes*. For Con_{PA} is true in *ML* and false in *ML**. With few exceptions everyone who understands what it means either accepts Con_{PA} or is neutral about its truth value. So by Ricketts' criterion *ML** is not compatible with our dispositions.

This argument does not go far enough to help Carnap, however, for at least three reasons. Firstly, one must keep in mind that Con_{PA} is just one example of an undecidable sentence, even though of course an especially interesting one. But there are *infinitely many* other purely mathematical sentences which are independent of the axioms of PA, and for most of them we will have no inclination to either assert or deny them. So even if our speech behaviour *excludes* the particular metalanguage *ML** Beth describes, there will still be infinitely many alternative metalanguages *ML***, *ML****, etc., that are compatible with it, and which differ concerning the truth value of some undecidable sentence. Using Ricketts' criterion, one can thus at most argue that we do *not* speak *ML**, but we cannot draw the further conclusion that we *do* speak *ML*, rather than one of the infinitely many other non-standard metalanguages.

Secondly, consider the following hypothetical scenario: as a matter of fact Goldbach's conjecture can be derived from the Peano axioms, but the mathematical public has been convinced by a proof with an as yet undiscovered mistake in it that the negation of Goldbach's conjecture is provable in PA. In this case, should we conclude that PA is not an adequate way to capture the language of these mathematicians, since they a sentence that is false in PA? That does not seem right. What the mathematicians mistakenly believe is that their dispositions to reason in line with the Peano axioms commit them to also accept the negation of the Goldbach conjecture. The latter disposition is not supposed to be a wholly independent *addition* to these more basic dispositions, but rather a *consequence* of them. If this line of thought is correct, then Carnap would at least have to argue that our acceptance of Con_{PA} is not a brute fact, but somehow flows from the more basic dispositions that constitute our acceptance of the axioms of PA. And it is hard to see how this could be done, given that the axioms corresponding to the basic dispositions don't entail Con_{PA}.[9]

Thirdly, there are further problems for Carnap if we look at theories stronger than Peano arithmetic. It is indeed plausible that we in fact accept Con_{PA}, since the consistency of PA is regarded as well-established. But the case of PA was just one convenient example, and, in light of the principle of tolerance, we can also use and discuss various other logical systems, such as Zermelo-Fraenckel set theory (ZFC). If Carnap wants to treat ZFC in analogy to the case of PA, he needs to claim that we use a metalanguage which settles the consistency sentence Con_{ZFC}, and in general for any theory T we want to use, we would need to speak a metalanguage in which Con_T is true. But for many logical systems, such as ZFC plus some large cardinal axiom, it will be an open question whether the system is consistent or not, and hence there will be no established disposition to either assert or deny the relevant consistency sentence. So even if we grant that we speak a metalanguage that settles the consistency claim for the specific case of PA, this strategy does not generalise in the way Carnap's overall philosophy of mathematics requires.[10]

Let me sum up my interpretation of Beth's objection. Carnap's broadly behaviourist conception of the relationship between natural and formal languages can accommodate the using and sharing of *weak* metalanguages with recursive axioms. But, as we saw, in *Logical Syntax* Carnap uses a strong metalanguage which includes infinitary inference rules. And for metalanguages of this kind Carnap does not in general seem to be in a position to say that we have adopted one – such as *ML* – rather than another – such as *ML** – since our linguistic dispositions cannot single out one over the other. Consequently no sense can be made of *sharing* such a metalanguage either – and hence Carnap's reply to Beth fails to address the real problem.

The argument demonstrates that there is a tension between the two distinct

[9] This kind of worry is similar to issues that arise in discussions of what has been called the *normatively of meaning*. It seems that when we understand the word 'red', for instance, there is a sense in which we *should* only apply the word to red things and not to non-red things, regardless of whether we actually are disposed to do so. Kripke has forcefully argued that it is not easy to account for the relevant normativity (Kripke 1982), however. See Boghossian 1989 for a useful survey.

[10] Some general remarks on the relationship between theories and their consistency sentences are in order here. In the main text I frequently use Con_{PA} as an example of an undecidable sentence, since it is indeed independent from PA (as is Con_{ZFC} from ZFC). It is not the case, however, that for every theory T its consistency sentence Con_T is independent of it. The theory that results from adding $\neg\mathrm{Con}_{PA}$ to PA, for instance, is such that it proves the negation of its own consistency sentence, i.e. $\neg\mathrm{Con}_{PA+\neg Con_{PA}}$. It is hard to imagine anyone seriously using a theories of this kind, however, and I will therefore set them aside. See Warren 2015a for a scenario in which Martians reason in accordance with PA+$\neg\mathrm{Con}_{PA}$ though.

roles Carnap has for metalanguages. In the first instance they are needed to make communication possible:

> Means of Communication
> Unless participants in a discussion use the same metalanguage, there is no genuine communication between them.

And it is hard to deny that there must be some language which plays this role. What is distinctive of Carnap's philosophy of mathematics, however, is that he thinks that metalanguages also play a second role:

> Resolves Incompleteness
> Since by Gödel's incompleteness theorems no recursive mathematical theory is complete, a strong metalanguage with infinitary rules needs to settle the truth and falsity of undecidable sentences.

This move is supposed to capture the idea that mathematical truth is a matter of linguistic rules. Carnap seems to have assumed that relying on strong metalanguages for this task is unproblematic, since objections to their use would have to rely on philosophical considerations, and are hence excluded by the principle of tolerance. But Beth's argument shows that this is too quick. For when a formal language is actually to be *used* as a metalanguage, there needs to be a sense in which our linguistic dispositions correspond to the rules of this language. And this requirement severely limits what languages can be employed in this role.

If this is correct, Carnap thus faces a dilemma: resolving incompleteness requires strong metalanguages, using a metalanguage to communicate requires weak metalanguages.

4.2.2 Ontology and Ontological Commitments

In addition to the principle of tolerance, Beth also raised an objection against Carnap's attitude towards ontology:

(B_3) Carnap cannot treat all questions concerning ontology as a merely pragmatic matter (Beth 1963: 500f).

This point still requires a more detailed discussion, for it is not immediately clear how the considerations about metalanguages, infinitary rules, and dispositions that have occupied us so far are related to questions concerning ontology. It is the aim of this and the next section to elucidate the matter, and to argue that

Beth's case for (B_3) is strong as well. The overall idea is as follows: Carnap cannot account account for determinate facts about *syntax* without relying on ontological posits, understood in an external sense.

The following is probably the most explicit passage on this topic by Beth himself:

> And finally the notion of a standard model, if understood in accordance with strict usage, involves an appeal to certain ontological commitments. Therefore, in connection with the problem of the method of semantics, a discussion on the acceptance or rejection of certain ontological commitments cannot be avoided. And such a discussion cannot be restricted, as *Logical Syntax* suggests, to the question whether a certain phrase "for all properties ..." can be formulated in *S* (or M_2); for it remains to be seen whether this phrase "for all properties ..." can be interpreted as "for all properties whatsoever," and, if understood in accordance with strict usage, such an interpretation is impossible without an appeal to certain ontological commitments. (Beth 1963: 501)

Beth seems to be saying that we cannot use Carnapian metalanguages in accordance with strict usage without appealing to ontological considerations. But unfortunately it remains unclear what exactly such an appeal to ontology amounts to, and why this alleged connection between strict usage and ontology obtains. To make matters worse, some of Beth's remarks about the difference between the ontological commitments of Carnap and Carnap* are apt to confuse the reader:

> In particular, Carnap* will not accept the existence of our set of symbols:
>
> $$0, 0', 0'', \ldots \text{ etc.}$$
>
> by which we mean, the real set, without the additional elements which Carnap* wrongly supposes to be contained in it. On the other hand, we accept the existence of Carnap*'s so-called set of numerals, but we know that, in addition to all numerals it must contain still other elements. We even understand why Carnap*, because of his fatal inclination for nonnormal ontological commitments, is unable to see his own errors. As Spinoza said: *veritas norma sui et falsi est.* (Beth 1963: 499f)

Since by construction Carnap* accepts the existence of *more* numerals than Carnap does, it is peculiar that Beth now claims that Carnap* rejects the existence of the set of standard numerals, for this is just a *subset* of the set of numerals he believes in.

There is a way to make sense of these remarks however, based on a technical result known as the *overspill lemma* (Kaye 1991: section 6.1). The lemma shows that in PA it is impossible to define a formula $\phi(x)$ such that, if PA is interpreted in a non-standard model, $\phi(x)$ only applies to the standard elements. Instead some non-standard elements always spill over into the extension of $\phi(x)$. This result can be glossed as showing that it is impossible to characterise the standard numbers from within PA. And it seems reasonable to interpret Beth's claim that Carnap* is "unable to see his own errors" as alluding to precisely this fact: from his perspective it is impossible to pick out the set of numerals Carnap considers to be real – the standard ones, that is – and so there is some sense in which Carnap* cannot accept the existence of these numerals.[11]

In itself this observation does not suffice to explain the alleged problem with Carnap's purely internal treatment of ontology, however. I think that the source of the confusion is that Beth says too little about what he means by ontological commitment (which is also what Carnap complains about in his reply). For this notion was of course introduced and popularised by Quine, and so we need to be careful when using it in a discussion of Carnap. Here is one of Quine's well-known definitions:

> [A] theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true. (Quine 1953: 13f)

In section 1.1.1 we already saw that Carnap was not happy with the terminology Quine recommends, for he thought that using the name 'ontology' makes Quine's project appear too similar to traditional kinds of metaphysics they both reject. But, importantly, Carnap agreed with Quine that the connection between quantification and what exists is of great importance:

[11] The overspill lemma is frequently attributed to Abraham Robinson, but people rarely cite an exact source. A lot of his work on non-standard models was done in the 1960s, however, see for instance Robinson 1961: section 3.1 for discussion of something akin to the overspill result, even though with a different terminology. This makes my proposed reading of Beth somewhat anachronistic, since even though it only appeared in 1963, his paper must have been written in the mid to late 1950s. This can be deduced from the fact that in a footnote of his reply to Beth, Carnap has a note added in 1962 which refers to two books from 1958 and 1959 that have appeared "in the meantime" (Carnap 1963b: 927fn24).

> W.V. Quine was the first to recognize the importance of the introduction of variables as indicating the acceptance of entities. "The ontology to which one's use of language commits him comprises simply the objects that he treats as falling . . . within the range of values of his variables." (Carnap 1956a: 214n3)

In light of this Beth's claim that Carnap needs to accept certain ontological commitments seems correct but uninteresting. For if to be ontologically committed to numbers, for instance, is just to use a theory in which numbers are being quantified over, then Carnap of course agrees that we can and should use a framework which is ontologically committed in this way. In fact the whole point of Carnap's internal Platonism was to demonstrate that there is nothing objectionably metaphysical about ontological commitments understood in this sense. It is therefore understandable that in his reply to Beth, Carnap refers the reader to his distinction between internal and external existence questions (Carnap 1963b: 933).

I think that the point Beth intends to make is different, however. As I understand him, he wants to argue that Carnap needs to allow for cases where what exists is *not* just a purely framework-internal matter. In other words, Beth thinks that the case of Carnap* shows that there must be some *external* questions about ontology which *cannot* be treated as merely pragmatic.

In order to appreciate this argument, let us recall the overall strategy behind Carnap's internal Platonism. The idea was to make talking about abstract objects unproblematic by construing the relevant area of discourse as internal to a linguistic framework and analytic. One important fact that I didn't dwell on earlier, but which has become especially salient in the meantime, is that syntactic entities such as numerals, sentences, and proofs – and ultimately also the frameworks themselves – are abstract entities too. As Friedman usefully sums up, this is one of the important upshots of Beth's Carnap* case:

> The crux of Beth's argument is that syntax is itself a kind of arithmetic (as becomes especially clear in a Gödel numbering, for example). And, viewed as an arithmetic, a Carnapian syntax language or metalanguage may then have non-standard models – containing non-finite numbers (non-finite sequences of expressions) beyond the standard numbers 0, 1, 2, . . . (so that, in the case of syntax, there may be more than a finite number of numerals 0, $0'$, $0''$,..., for example, or derivations may have more than a finite number of steps). (Friedman 2009: 238)

I think that the best way to appreciate Beth's argument for ontology in an external sense is to consider how Carnap's internalist strategy can be applied to syntax, which is what I will do in the rest of this chapter.

Before I move on I want to briefly reflect on the abstractness of syntax. From time to time there have been attempts to construe syntactic entities as concrete, such as that explored in Nelson Goodman and Quine's well-known paper "Steps Towards a Constructive Nominalism" (Goodman and Quine 1947). Carnap himself was not uninterested in such radical forms of nominalism and finitism, and he discussed this issue with Quine and Tarski in 1941. An obvious problem quickly arises, however: for many purposes we need syntactic objects of arbitrary length, but identifying syntactic items with concrete entities threatens to impose arbitrary upper bounds. Carnap makes this vivid using an example:

> S_1 is proved through a proof which more-or-less fills up the largest star; further, a derivation of S_2 from S_1 more-or-less fills up the same star. But the concatenation of both chains of sentences is nowhere. Consequently, according to the proposed finitistic concepts, we cannot say that we have proved S_2. But every logician will surely want to say that if S_1 is proven and S_2 is derived from S_1, then S_2 is also proved (not merely "provable," which is inexpressible in this language). (Carnap in Frost-Arnold 2013: 165)

Carnap seriously considered whether we can make sense of the idea that even in a finite universe are there infinitely many *arrangement-possibilities* of the finitely many things. He sketches this approach as follows:

> I : If we have only finitely many things, and thus finitely many names 'a,' 'b,'...'Q,' then we can build arbitrarily long sequences:
>
> R(a, a)
> S(a, a, a)
> T(a, a, a, a)
> ...
>
> Naturally, in the same world, we cannot write down arbitrarily long sequences; but with the help of abbreviations, we can indeed talk about them. With these, we can build an unrestricted arithmetic. (Frost-Arnold 2013: 163)

The idea that there is a *potential* infinity of such sequences in the absence of any *actual* infinity of things is nothing new. In fact Hilbert's finitism was based on such a potential infinity of strokes (|, ||, |||, ...), as Bernays helpfully summarises:

> [...] the objects of intuitive number theory, the number signs, are, according to Hilbert, also not "created by thought". But this does not mean that they exist independently of their *intuitive construction*, to use the Kantian term that is quite appropriate here. But the construction always only yields either a single determinate figure or a procedure for obtaining a further figure from a given one (e.g. by affixing "+1"). But it does not lead to the idea of a simultaneous existence of "all" the number signs. (Bernays 1998: 226).

It seems that in the discussion at Harvard, Quine and Tarski – who were unsympathetic towards modal notions and the distinction between potential and actual infinites anyway – convinced Carnap that this approach is unfruitful.[12] While the consequence they drew was to take a closer look at nominalism, Carnap eventually concluded that the antipathy towards accepting the existence of infinitely many objects was based on a confusion. We saw that by the time he wrote "Empiricism, Semantics, and Ontology", Carnap had come to accept what I called internal Platonism, according to which nominalist scruples are unmotivated. In the following I will therefore focus on how the mature Carnap can account for syntax, and set the earlier attempts to identify syntactic entities with anything concrete aside.

4.2.3 The Determinacy of Syntax

Let us once again consider the theory of Peano arithmetic. Here are some properties PA has:

> '2+2=4' is derivable from PA.
>
> 'There are infinitely many primes' is derivable from PA.
>
> '0=1' is not derivable from PA.
>
> PA is consistent.

PA is a formal theory, and hence an abstract syntactic object. Consequently the properties I just mentioned are properties of an abstract object. And this means

12 I will come back to the question whether modality can help Carnap in section 7.2, and argue for a negative answer.

that, if Carnap's internalist treatment of all abstract objects is to succeed, he needs to account for these properties by virtue of linguistic frameworks somehow.

As I understand him, Beth thinks that the case of Carnap* demonstrates that we run into problems when accounting for syntax in a way compatible with internal Platonism. For using Gödelisation we can encode claims about what PA does or does not entail in the language of arithmetic, and hence express them within PA. And it appears to be a consequence of this that, depending on what model we use to interpret PA, we can either regard it as consistent or inconsistent:

> This remark implies at once a positive solution to a problem which I stated once in connection with a critical analysis of *Logical Syntax*: do the syntactical properties of a given object language O depend on the choice of a syntax language S? For, from our point of view, the choice of S allows us to believe in the inconsistency of II, the choice of S* allows us to prove the inconsistency of II, whereas the choice of a suitable extension of II allows us to prove the consistency of II. (Beth 1963: 478)

This point can be made more clearly if one keeps in mind the distinction between linguistic frameworks, which comprise both an object and a metalanguage, and the mere object language PA. The internalist would have no problems with syntax at all if any arithmetised claim about syntax would just be straightforwardly derivable from PA. But by Gödel's second incompleteness theorem we know that this is not so, with Con_{PA} being the salient example. For this reason it is not possible to account for all the syntactic facts about PA by adopting the following:

Syntactic Proposal
ϕ is true **iff** PA $\vdash \phi$.

Carnap's way out was to conceive of frameworks as bundles of object and strong metalanguages, which jointly determine truth and falsity for mathematical sentences. We thus ended up with a linguistic framework that combined the object language PA with a metalanguage which referred to the standard model of arithmetic, resulting in the following definition:

Semantic Proposal
ϕ is true **iff** $\mathbb{N} \vDash \phi$.

One might think that this solves the problem, for since the standard model $\mathbb{N}$ settles the truth value of every arithmetical claim, it also settles the status of all claims about syntax. The problem with this, however, is that one can construct a non-standard model of PA M in which Con_{PA} comes out false:

> Competing Semantic Proposal
> ϕ is true **iff** $M \models \phi$.

Since, as Beth stresses in the quote, the object language PA does not uniquely determine one of these competing interpretations, we thus end up with two distinct linguistic frameworks: one based on the Semantic Proposal, the other on the Competing version. And so the internalist who wants to account for syntax has to deal with an embarrassment of riches: if we want to stick to the view that all syntactic claims have one determinate truth value, we need some reason to discount the deviant framework based on the Competing Semantic Proposal as being irrelevant for matters of syntax.

Based on this observation, we can begin to understand why Beth thinks that there must be some ontological questions that are both factual and external. We get there by combining the following two theses:

> Determination Thesis
> If the axioms of some object-language *OL* leave the truth value of some sentence unsettled, and there are metalanguages M_1 and M_2 that determine different truth values for this sentence, then the sentence simpliciter has no determinate truth value – *unless* some non-linguistic factor settles the matter.

> Determinate Consistency
> The syntactic properties of object-languages such as PA, in particular their consistency, are *not* indeterminate.

That Beth holds the Determination Thesis was already suggested by the quote above. It is further supported by the fact that Beth cites a review of *Logical Syntax* by Kleene, who reasons as follows:

> If logic and mathematics are taken as wholly formal, one apparently must reject the conception, that a sentence S of classical mathematics is necessarily either true or false in such a sense that the problem, to determine which is the case, belongs to logic and mathematics when S (or "S is analytic") is irresoluble in terms of d-rules already stated. (Kleene 1939: 84f)

The d-rules are the purely syntactic rules of derivation in the object language, and hence Kleene's conclusion is in effect that mathematical sentences which are undecidable in PA have no determinate truth value.

Beth doesn't state the Determinate Consistency assumption explicitly anywhere, but we can be pretty sure that he would have agreed to it. For even a conventionalist and proponent of tolerance, who thinks that we are free to adopt any system of rules and axioms we like, will presumably say that *once the rules are accepted* it is not a further matter of choice what follows from them. We in fact already saw some evidence that Carnap endorses such a position in section 1.2.3, and Gödel also considers this assumption to be an obvious truism:

> Some body of unconditional mathematical truth must be acknowledged, because even if mathematics is to be interpreted to be a hypothetico-deductive system, still the proposition which states that the axioms imply the theorems must be unconditionally true. [...] For while the definitions of 5, 7, 12 and the rules of computation for + and = [...] seemingly can be interpreted as conventions, the statement that 5+7=12 follows from these conventions evidently expresses an objective (combinatorial) fact (Gödel 1995b: 200).

In chapter 6 we will encounter a form of radical conventionalism that denies this assumption, but I will not question it for now.

The argument for external ontology now proceeds as follows: if we want determinate consistency we need some language-external factor that settles the truth values of sentences that are left indeterminate by the purely linguistic rules, such as Con_{PA}. And the obvious candidate to play this role would be something like the Model-in-the-Sky of the external Platonist. If we had such a thing at our disposal, we could set the aside the Competing Semantic Proposal on the grounds that it misrepresents the realm of arithmetic, since it is committed to non-standard syntactic entities to which nothing in the *real* model corresponds. The price of this move, however, is to introduce a framework-external sense of existence, and hence to give up on full-blown internalism. There needs to be an exception at least for syntax.

Against this one might object that the Determination Thesis is false since it relies on an implausible picture of how object languages are related to metalanguages. It is of course true that the object language of PA does not *in itself* somehow force us to adopt a metalanguage which settles the truth values of undecidable sentences in one way rather than the other, otherwise the incompleteness theorems would not be a challenge. But this result does not entail that all *possible* metalanguages are on a par. For one must *describe* an object language from the perspective of a particular metalanguage after all, and so one possible metalanguage is therefore privileged in virtue of being the one that is actually *used*. There thus seems to be possibility Beth has overlooked, namely a way to single out one of the many possible metalanguages without relying on external ontological posits.

This is in fact how I understand Carnap's own reply to Beth: he thinks that we actually occupy the position of someone who uses metalanguage *ML*, which encapsulates standard syntax, and can hence dismiss Carnap*'s conception of syntax as non-standard and unintended without relying on any wholly language-transcendent notions. But here my objection to Carnap's *we in fact speak ML* response from section 4.2.1 strikes again. It is of course fine to maintain that we always speak from the perspective of one metalanguage or another. But what we need here is a metalanguage that settles the truth values of all undecidable sentences of PA, and hence the argument from dispositions against non-recursive inference rules applies. This being the case, Beth's claim that only external ontological assumptions can do the trick is actually well-supported. A number of further attempts to avoid this conclusion will be discussed in chapter 7, but before that I will address some exegetical questions.

Chapter 5

Alternative Interpretations of Beth

Beth's argument from non-standard models has received some discussion in the secondary literature, and a number of scholars have defended Carnap against Beth's objection. In this chapter I survey this literature and argue for the following two points: first, extant responses to Beth on behalf of Carnap cannot be applied to defuse my interpretation of the argument, since it differs from those others have given in important ways. Secondly, although the exegetical situation is murky at times, there are good reasons to think that my interpretation also better fits the argument Beth himself had in mind than competing readings.

In the first half of this chapter I compare Beth's argument to other well-known arguments that rely on model-theoretic considerations, namely Putnam's argument against metaphysical realism and Skolem's paradox. I show that it is tempting to read Beth's argument as a version of these established argument, but argue that the temptation should be resisted. In the second half I consider attempts to attack Beth by arguing that he misrepresents the nature of Carnap's position. I show that these proposals are either unconvincing or do not suffice to solve the underlying problem.

5.1 Beth and Model-Theoretic Arguments

5.1.1 Non-Standard Interpretations

There are a number of influential arguments based on model theory in philosophy, and it is therefore natural to wonder whether Beth's argument is in effect a version of a more established one. The two most salient candidates for comparison are Putnam's argument against metaphysical realism, and Skolem's paradox about the relativity of set-theoretic notions. In this and the next few sections I will argue that Beth's argument should *not* be identified with either of these, which also makes some defences of Carnap in the secondary literature ineffective. It will be convenient to start the discussion by introducing some terminology to distinguish between different kinds of non-standard models.

Using PA as an illustration, the first kind of non-standard interpretation of PA is one where the domain is isomorphic to the standard model $\mathbb{N}$ but contains different elements. Since there is a one-one mapping between the natural numbers and the even numbers, for instance, a model in which the domain contains only even numbers can make the same sentences true as the standard model. Non-standard models of this kind are called *isomorphic*:

> Isomorphic
> A model which is isomorphic but not identical to the standard model.
> *Example:* Interpreting PA in the sets of even numbers by mapping each number n to $2n$.

The second kind of non-standard interpretation involves models which are not isomorphic to the standard model. By the upward Löwenheim-Skolem theorem, for instance, we know that PA has some uncountable models. Since the natural numbers are countable, such an uncountable model cannot be isomorphic to the standard model. In contrast to the third kind of non-standard model I will describe later, however, interpreting PA in such a model does not show up in the object-language, since every sentence of PA that is true in the standard model is also true in the non-standard model. This property is called *elementary equivalence*, and the second kind of non-standard model can thus be characterised as follows:

> Non-Isomorphic but Elementarily Equivalent
> A model which is not isomorphic to the standard model but makes the same sentences true as the standard model.
> *Example:* Interpreting PA in an uncountable model via the upward Löwenheim-Skolem theorem.

The third and last kind of non-standard model is the most interesting, since it involves changes of the truth values of certain object-language sentences. We have already seen that PA has non-standard models of this kind by considering Gödel's incompleteness theorems. Since the sentence Con_{PA} is independent of the axioms of PA, both PA+Con_{PA} and PA+$\neg\text{Con}_{PA}$ have models. In the standard model Con_{PA} is true, however, and there is thus a non-standard model which gives a sentence in the language of PA a different truth value compared to the standard model of arithmetic. I call these kinds of non-isomorphic non-standard models *truth-switching*:

> TRUTH-SWITCHING
> A model which is not isomorphic to the standard model and changes the truth value of some sentences from those of the standard model.
> *Example:* Interpreting PA in a model where $\neg \mathrm{Con}_{PA}$ holds.

My interpretation of Beth's argument in the previous chapter crucially relies on the possibility of truth-switching non-standard models – for otherwise Carnap and Carnap* could not disagree on the facts of syntax. Not all philosophical uses of non-standard models are like this, however, and it has for instance been argued that the availability of isomorphic non-standard models suffices to put some conceptions of the relationship between language and the world under pressure.

This is best illustrated by considering Putnam's model-theoretic argument against metaphysical realism. A lot could be said about what exactly is characteristic of the flavour of realism Putnam intends to attack, but for our purposes the following characterisation is a good starting point:

> *Metaphysical* realism, on the other hand, is less an empirical theory than a model – in the "colliding billiard balls" sense of 'model'. It is, or purports to be, a model of the relation of any correct theory to all or part of THE WORLD. [...] In its primitive form, there is a relation between language and a piece of THE WORLD (or a kind of term). (Putnam 1977: 483f)

Talking about 'THE WORLD' in capital letters here is supposed to convey the idea that reality is in itself divided into objects, wholly independently of our ways of conceptualising it. The relationship between the world and our best theory can then be thought of as follows: the objects making up reality are the *intended interpretation* of the theory, and, 'there are cats' is a true sentence of our theory if and only if some of the objects reality consists of are indeed cats.

The underlying idea that there is some kind of correspondence between reality and our theories is a natural one, but Putnam thinks that it faces some deep problems. One of his objections – the argument from completeness – begins as follows:

> Since the ideal theory T_1 must, whatever other properties it may or may not have, have the property of being *consistent*, it follows from the Gödel Completeness Theorem [...] that T_1 has models. (Putnam 1980: 473)

More specifically, one can show that every consistent first-order theory has a model *in the natural numbers*. And this seems to be an uncomfortable result for the metaphysical realist. For whatever we may think the intended interpretation of our best theory – which is just THE WORLD itself – looks like exactly, it seems unlikely that it is purely mathematical. If this is so, however, the metaphysical realist needs to answer the following question: why exactly should we suppose that the model of our best theory is the intended interpretation rather than some purely mathematical model, given that both make the very same sentences of the theory itself true?

I will not go further into Putnam's reasons for thinking that the metaphysical realist has no good answer here. For whether successful or not, it should hopefully be uncontroversial that Carnap is *not* a metaphysical realist, and is hence not a target of Putnam's attack. In the earlier discussion of Carnap's treatment of the reference-relation (section 1.1.3) it already emerged that he has in effect a disquotational conception of reference: if within a framework we can express that 'cat' refers to cats then, for Carnap, there is no further question of whether 'cat' might *really* refer to numbers after all. In in a later section we will also see that, unlike a metaphysical realist, Carnap at times actually endorsed the idea that all our quantifiers range over sets and numbers (section 9.1).

I think that, for this reason, there is no way to argue against Carnap's position based purely on the existence of isomorphic non-standard models. Putnam's argument attacks a way of thinking about the relationship between theories and reality that Carnap would reject anyway. This result will probably not come as a huge surprise, as it is not very tempting to read Carnap as a metaphysical realist in the first place. We can therefore move on to an interpretation of Beth's argument that is much more plausible but still distinct from mine: namely reading Beth as putting forward a version of Skolem's paradox, which relies on the possibility of non-isomorphic but elementarily equivalent non-standard models.

5.1.2 Ricketts on Beth and Skolem

Right after describing Carnap*, Beth writes that "the above considerations [...] are only variants of the Löwenheim-Skolem paradox" (Beth 1963: 478), an influential argument first presented by Skolem in 1922 that is nowadays usually just called *Skolem's paradox* (Skolem 1967) . Ricketts has taken this reference very seriously, and interprets Beth's whole argument as a version of this (alleged) paradox. From this he then concludes that Beth poses no serious threat to Carnap's position, which is not surprising, since Skolem's paradox is also not usually regarded as a genuine paradox.[1] I will first introduce Skolem's paradox, describe Ricketts' reconstruction of Beth, and then show that it fails to capture the actual argument.

One consequence of the Löwenheim-Skolem theorem is the following:

> If a first-order theory has an infinite model, it has both countable and uncountable models.

As yet there is nothing paradoxical about this, but things get interesting if we apply this result to a first-order axiomatisation of set theory, such as ZFC. For in ZFC we can express the distinction between countable and uncountable sets, and, furthermore, can prove the following:

> There exists an uncountable set, i.e. a set which is not isomorphic to the countable natural numbers.

By the Löwenheim-Skolem theorem ZFC has a countable model – i.e. it can be interpreted in a domain that is isomorphic to the natural numbers. And one might think that this result is contradictory. For how could ZFC be true in a countable model even though it proves the existence of *un*countable sets?

This tension is merely apparent though. For the provability of the existence of an uncountable set within ZFC amounts to the following: in every interpretation of ZFC there are two infinite sets such that there is no one-one mapping between them. Looked at from the outside, in a countable model both such sets will be countable as well, provided that the model is transitive.[2] One of them will nevertheless *seem* uncountable from *within* the theory ZFC, and this is because the interpretation must be such that it fails to contain the function that would map the objects from both sets to each other. To put this differently: when we say that there is no one-one function between sets, we usually mean to quantify over *all functions whatsoever*. But if ZFC is interpreted with respect to a specific

1 See Bays 2014 for an overview of the most common philosophical responses.

2 A transitive model is such that if its domain contains a set, then all members of this set are also members of the domain. A non-transitive model, on the other hand, can have countably many objects in its domain, while one of these objects is an uncountable set such as the set of real numbers.

model, the quantifiers in ZFC merely range over the functions contained in that model. And so although two sets will be countable for someone who quantifies over all functions, one of them can appear to be uncountable to someone who quantifies over fewer functions.

In a nutshell, the apparent paradox is resolved as follows: if interpreted in a countable model, the sentence of ZFC that apparently asserts the existence of an uncountable set doesn't *really* express this claim after all. This is because, from our perspective, we can see that the quantifiers of this interpretation leave out some functions that are relevant to the question of countability. There is hence no genuine paradox, for there is no one claim – such as *there are uncountable sets* – for which we have reasons for and against accepting it. Instead we can always use Cantor's theorem to show that there are uncountable sets, and while it might be that sets that seem uncountable from one perspective appear countable from another standpoint, this is no genuine contradiction because ″uncountable″ has different meanings in the two perspectives, thanks to the difference in what functions are being quantified over.

One might nevertheless be tempted to think that, even though the original paradox has been defused, some reason to be concerned remains. Thomas Tymoczko gives voice to the nagging doubts someone might have as follows:

> Maybe the reals are *really* countable and we simply lack the 1-1 function that shows this. Maybe we are living in someone else's countable model! (Tymoczko 1989: 289)

He immediately points out, however, that on reflection this worry is just confused:

> There is something crazy about this suggestion. After all, we can prove that the reals are uncountable. If Cantor's mathematical proof is not good enough to settle the matter, what would settle it? Surely it doesn't make sense to wonder whether the reals are ′really countable′ according to some hypothetical alien concept of countablity to which we, by hypothesis, have no access. [...] We construct the alternative [countable model] by leaving certain things out of the model (e.g., almost all the real numbers). It is totally incoherent to turn around and wonder whether *that* world might not be ours after all. (Tymoczko 1989: 289f)

And for Carnap this is clearly the right thing to say, given that he is strictly opposed to any completely language- and theory-independent external perspective. Suppose we adopt a framework in which we can prove that there are uncountable sets, and meet someone with a different framework from whose perspective our sets will seem countable. We can then discuss whether it might be *convenient* to adopt this other, apparently more powerful, framework. But for Carnap there can be no sense in which we would be making a factual mistake if we just stick with the framework we have.

I therefore think that one cannot generate trouble for Carnap's position by purely relying on Skolem's paradox, which is why I stressed the importance of Gödel's incompleteness theorems in Beth's argument. Let us now look at how Ricketts reads Beth:

> In his reply to Beth, Carnap observes that Beth's point really turns on the possibility of a tacit divergence between two logicians in their understanding of the informal syntax language they use to state the transformation rules for Language II. One logician might understand the arithmetic in the informal syntax language standardly; the other might understand it non-standardly. *This divergence need not be manifest in their use of the sentences of the informal syntax language.* [...] Having read our Skolem, we observe that we can state transformation rules in two different ways, one corresponding to the standard model of arithmetic, another corresponding to a non-standard model. *We may suppose that each group of transformation rules demarcates the same formulas of the object calculus as true.* (Ricketts 2004: 194, my emphasis)

As I understand this, the scenario Ricketts describes here is not like the case of Carnap*. For Carnap and Carnap* *do* manifest divergent usage in the informal syntax language, and do *not* agree on the truth values of all object language sentences: they have incompatible views about the consistency of PA, for instance. What Ricketts has in mind is rather a character we may call Carnap**, who like Carnap* reads *Logical Syntax* and interprets in a non-standard way, but generates his non-standard interpretation by applying the Löwenheim-Skolem theorem rather than Gödel's incompleteness theorems. Carnap** then uses a metalanguage with a non-standard interpretation, but one that is elementarily equivalent to the standard interpretation. He and Carnap thus agree on the truth values of all sentences, and so for this scenario Ricketts' description fits.

Ricketts concludes that the possibility of Carnap** is no problem for Carnap, and with this I agree. For Carnap** would only be troublesome if it would make sense to worry about the possibility of 'inhabiting' some non-standard model without this showing up in the truth values of any sentences, just like we saw with respect to ZFC and uncountable sets. And this is clearly not an issue that Carnap should take seriously.

Supposing that I am right, my reconstruction of Beth's argument from the previous chapter is therefore damaging to Carnap's position in a way that the one discussed here is not. One may nevertheless worry that in several ways the actual text of Beth's paper fits Ricketts' version much better than mine. For one thing, I quoted Beth's reference to Skolem's paradox at the beginning of the section. For another, in my reconstruction *linguistic dispositions* play a major role, which are not something Beth explicitly writes about at all. It is therefore natural to wonder whether what I have defended is *my own* argument from non-standard models rather than *Beth's*. This exegetical question is maybe less important than the systematic question of whether the argument succeeds, but in the following I nevertheless want to defend my ascription of the argument to Beth.

5.1.3 Exegetical Traps and Pitfalls

It is not easy to interpret Beth's paper in a way that perfectly fits all of the claims he makes. Some initially confusing remarks become intelligible once one appreciates that the scenario Beth describes is easily misinterpreted. It is for instance very tempting to describe Carnap and Carnap* as disagreeing about a common subject matter, namely the object language of Peano arithmetic, but this is not really appropriate since they do not actually share a metalanguage. This is a point Carnap stresses as well:

> Therefore it seems to me misleading to say that Carnap* has views about the languages II and II* which diverge from our views about these languages. It seems to me more correct to describe the situation as follows: (a) Carnap* does not use the metalanguage *ML*, but a language *ML** which, although it uses the same words and sentences, differs from *ML*, since some of the words and sentences have different meanings; and (b) since the labels "II" and "II*" have in *ML** meanings different from those in *ML*, Carnap* is not talking about the same languages as Carnap. (Carnap 1963b: 928f)

For expository purposes it is nevertheless helpful to conceive of Carnap and Carnap* as looking at one and the same object language, since otherwise it is hard to understand what Carnap*'s non-standard interpretation is like in the first place. But this description only makes sense from *our* perspective, assuming that we are like Carnap and unlike Carnap*. Someone who was actually like Carnap*, on the other hand, would not have been able to follow the construction of a non-standard interpretation by adding $\neg\text{Con}_{PA}$ at all, since for them neither PA nor any extension of it would have *any* model due to their inconsistency.

Some of Beth's remarks remain elusive, however, even if one keeps the distinction between Carnap and Carnap*'s perspectives firmly in mind. In a footnote he for instance suggests that for Carnap* languages II and II* have no finite axiomatisation (Beth 1963: 478n27). But on my reading this is clearly false, since for Carnap* both of these theories are inconsistent, and finitely axiomatising the inconsistent theory is easy. I can therefore not completely rule out the possibility that Beth had a slightly different scenario in mind after all, and that another interpretation could be given. I think that there are good reasons to just dismiss the remark from the footnote as a simple mistake on Beth's part, however. For as we will see now, Beth makes at least one other claim which is clearly a factual mistake.

As it happens, the relevant mistake is precisely the claim which made Ricketts's interpretation compelling: namely that "the above considerations [...] are only variants of the Löwenheim-Skolem paradox" (Beth 1963: 478). As I already mentioned this comes right after the introduction of Carnap*, and hence Beth seems to say that the possibility of Carnap* follows from the Löwenheim-Skolem *theorem*. But that is clearly not true. Based on Löwenheim-Skolem, we can indeed show that theories such as PA or Carnap's language II have non-standard models, and even that they have what I called non-isomorphic but elementarily equivalent interpretation – so the possibility of Carnap** from the previous section does follow. But the essential feature of Carnap* was that he and Carnap come to different verdicts concerning the consistency sentence Con_{PA} – so the interpretations they use are not elementarily equivalent, but assign different truth values to some sentences. And in order to show that such *truth-switching* non-standard interpretations are possible, we actually need to rely on Gödel's incompleteness theorems, and not merely on the Löwenheim-Skolem theorem.

The point I just made would be moot if all theories that have non-isomorphic but elementarily equivalent non-standard models also have truth-switching non-standard models. But this is not so. Earlier I briefly mentioned the non-recursive theory True Arithmetic. By construction this theory is *complete*, since each arithmetical sentence is such that either it or its negation is an axiom, and so there can be no truth-switching interpretation. Nevertheless True Arithmetic has uncountable non-standard models that are not isomorphic to the standard model $\mathbb{N}$, and hence True Arithmetic is an example of a theory which has the one kind of non-standard models but not the other.

This observation hopefully undermines the suggestion that a reconstruction of Beth along Ricketts's lines is preferable as a matter of exegesis. But it doesn't really address the second worry I alluded to, namely that the argument I presented makes heavy use of considerations about language use and dispositions, which are themes that do not appear explicitly in Beth's paper at all. To put this in a slogan, one might object that my interpretation of Beth makes his argument insufficiently model-theoretic, and too much like a specific version of the rule-following problem. In the following I will respond to these concerns, although the discussion is much more speculative than the previous point.

The crucial idea of Beth's argument seems to be that formal languages do not uniquely determine an intended interpretation, which is demonstrated by the case of Carnap*, who learns the letter of *Logical Syntax* but misses its spirit by interpreting the book in an unintended way. If this is the core of Beth's point, however, then this problem should affect *all* theories that can be interpreted in more than one way – i.e. any consistent first-order theory. But it turns out that Beth's paper contains some remarks that are directly in conflict with this conclusion. They come towards the end of the paper where, after discussing other aspects of Carnap's position, Beth comes back to the case of Carnap*. Other commentators have not paid much attention to these passages so far, presumably because they have the appearance of afterthoughts that do not substantially affect Beth's main point. But as will show now, a close reading of them actually reveals some important clues about how Beth intended his argument to work.

Beth begins by clarifying the nature of his opposition to Carnap. He stresses that he is sympathetic towards the principle of tolerance, but sees the need to restrict its scope when a particular language is supposed to be used as a metalanguage. He then distinguishes between metalanguages of varying strength, and discusses the problems that arise:

> In the first place, such a language, which may be called M, should enable us, as pointed out by Church, to state the necessary directives for the concrete manipulation of certain physical objects, namely, the signs of the object language. This implies that M must contain the means of expression for a certain version of elementary arithmetic or of a suitable general arithmetic. Moreover, this part of M, which will be called M_1, must be *understood in accordance with strict usage*.

> This demand, however, strongly restricts the development of M_1, the language of elementary syntax, into a means of expression, called M_2, for theoretical syntax; this follows from our discussion in Section 6. (Beth 1963: 499f)

M_1 must be some elementary form of arithmetic, since it needs to be used to describe the manipulation of syntactical objects. I therefore assume that we can regard M_1 to be weak theory such as Robinson arithmetic. Furthermore I assume that by the language of theoretical syntax M_2, Beth has in mind something like Carnap's metalanguage of *Logical Syntax ML*. This is a much stronger language than M_1 partly because it includes the infinitary inference rules we have seen in action. It is striking, however, that in the quote Beth seems to say that there is *no* problem about using M_1 in accordance with strict usage. And this is confirmed by the following remark:

> If we do not wish to be caught in the same trap as Carnap*, and, for instance, to be compelled to conclude that certain consistent theories are inconsistent, then we ought to be cautious in carrying out the passage from M_1 to M_2. Hence the question arises as to which precautions we could take in this connection. Unfortunately, it is impossible to point out precautions which are fully adequate. (Beth 1963: 499f)

This is remarkable since it strongly suggests that the issue of strict usage does not reduce to whether a theory has non-standard models. For the second incompleteness theorem also holds for very weak theories of arithmetic, including Robinson arithmetic, which has a finite number of axioms.

One might suspect that Beth was simply unaware of these results, and mistakenly thought that incompleteness doesn't arise for elementary arithmetic. But that is not plausible, since this result was proved in the influential *Undecidable Theories* by Tarski, Mostowski and Robinson, and Beth was actually one of the proof readers of the book (Tarski et al. 1953: ix). Of course Beth might still have been confused in his argumentation here, but charity demands that we at least try to read him in a way that makes sense of these remarks. I therefore think that the right conclusion to draw is that Beth's initial description of what it means to understand a language in accordance with strict usage was misleading, or at least incomplete.

On my own interpretation of Beth's argument, these last passages are much less mysterious. For there is a clear sense in which, say, Robinson arithmetic is less problematic than Carnap's *ML*. Since Robinson arithmetic has finitely many axioms and recursive inference rules, there is no deep puzzle about how we could have dispositions corresponding to this theory. As I have argued at length already, this is not so for *ML*. I think that this fact provides indirect evidence for my hypothesis that considerations about linguistic dispositions are relevant to Beth's notion of 'strict usage' after all, even though some creative exegesis was required to tease this out.

In the end the systematic question of whether there is a successful argument against Carnap is more important though, and I will now leave guesses about what Beth's true intentions were behind. In the next section I will discuss and reject some more attempts to defend Carnap's position.

5.2 Other Approaches

5.2.1 Friedman for and against Beth

Since the late 1980s Michael Friedman has discussed Beth's argument in a number of papers. His opinions have changed considerably over the years, however. In "Logical Truth and Analyticity in Carnap's *Logical Syntax of Language*"[3] from 1988, he thinks that the considerations of Beth (and Gödel) show that Carnap's treatment of analyticity is flawed. In response to Goldfarb and Ricketts he modified his arguments, but in his "Tolerance and Analyticity in Carnap's Philosophy of Mathematics" from 1999 he still holds that Carnap's account faces a deep tension (Friedman 1999b). But by 2009 Friedman has renounced his earlier criticism, and now in effect argues that Carnap's response to Beth is all that needs to be said (Friedman 2009: 241).

The basic idea behind the earlier Friedman's arguments was that Carnap's employment of strong metalanguages – i.e. metalanguages that contain non-recursive rules – is philosophically dubious. This is of course a hunch many people shared, but we already saw that it is not so easy to actually turn it into a non-question begging argument against Carnap's view. I think that Friedman's two attempts are ultimately not successful as they stand. It is instructive to understand why this is so, however, as I think that the earlier Friedman's skepticism towards strong metalanguages is closer to the truth than his later move towards agreement with defenders of Carnap like Goldfarb and Ricketts.

[3] I am citing the reprint of this paper from Friedman's collection *Reconsidering Logical Positivism* as Friedman 1999a.

In the first version of his argument, Friedman directly attacks Carnap's use of non-recursive resources in his definition of analyticity, on the grounds that this move is incompatible with the proclaimed aim of using purely syntactic methods to account for mathematics:

> Here is where Gödel's Theorem strikes a fatal blow. For, as we have seen, Carnap's general notion of analytic-in-L is simply not definable in logical syntax so conceived, that is, conceived in the above "Wittgensteinian" fashion as concerned with the general combinatorial properties of any language whatsoever. *Analytic-in-L* fails to be captured in what Carnap calls the "combinatorial analysis ... of finite, discrete serial structures" (§2): that is, primitive recursive arithmetic. (Friedman 1999a: 176)

But this objection is too quick. It is certainly true that in §2 of *Logical Syntax* there are passages which sound as if Carnap conceives of syntax in the way Friedman describes. If we take these remarks seriously then Carnap's program would in effect be a version of Hilbertian finitism: some weak theory such as primitive recursive arithmetic is regarded as understood and philosophically unproblematic, and the task is to justify the use of the more powerful remainder of mathematics on this slender basis. In light of Gödelian incompleteness, Hilbert's program is nowadays widely regarded as a non-starter (Smith 2013: section 37.5).[4] Regardless of what we think about this, however, it is obvious that Carnap's way of defining analyticity, with its reliance on infinitary rules, does not fit the model of the Hilbertian finitist at all.

It therefore seems best to conclude that Carnap's project is different from Hilbert's, and that his identification of syntax with the "combinatorial analysis [...] of finite, discrete serial structures" (Carnap 1937a: 7) should be taken with a grain of salt. For later in the book Carnap is perfectly open and upfront about his use of non-recursive methods, and since he repeatedly discussed Hilbert's program with Gödel he can hardly have been unaware of these differences. Goldfarb and Ricketts have therefore criticised Friedman's first argument as attacking a straw man, and not the actual Carnap of *Logical Syntax* (Goldfarb and Ricketts 1992: 65f).

Friedman accepted that his initial objection construed Carnap in a too Hilbertian fashion, but he maintained that one can still show that the use of strong metalanguages is problematic. The second version of his argument is not as

4 See Detlefsen 1986 for a more optimistic perspective.

straightforward, but the problem Friedman sees arises in situations where someone who uses a strong metalanguage encounters someone who only accepts a weaker language. Since Carnap at one point considers such an example himself, let us start by introducing a concrete scenario: logician *More* accepts language L_{More} in which one can quantify over individuals, sets of individuals, and sets of sets of indidivuals, whereas logician *Less* accepts language L_{Less} whose quantifiers range over fewer entities, namely only individuals and sets of those individuals.[5] Carnap comments on this situation as follows:

> Both logicians understand the syntactical rules for both languages. Both are in agreement with respect to many results concerning the two languages, in particular, with respect to syntactical results. Although [*Less*] understands neither language [L_{More}] nor its semantical rules, he can nevertheless learn to manipulate the sentences of this language according to the syntactical rules, and even to manipulate the semantical rules and semantical statements about [L_{More}], if they are stated in the form of a semantical axiom system whose syntactical metalanguage he understands. (Carnap 1963b: 873)

Less and More can discuss the merely syntactic properties of their languages without any problems. But there is an important asymmetry between them. More interprets Less's claims by disquotationally translating them into his own language, whereas Less is not in a position to proceed in the same way, because in his language "there is the set of sets of apples" is either meaningless or always false, regardless of the truth value it has for More. It is in this sense that Less cannot understand More's language in the same way More can understand Less's.

As I understand Friedman's argument, he thinks that this scenario is incompatible with the principle of tolerance. He considers the analogous case of the encounter of an intuitionist and a classical mathematician, and writes that the "Carnapian proponent of classical mathematics [...] can show that the mere idea that classical mathematics is analytic itself rules the intuitionist out of court" (Friedman 1999b: 230) – which in turn is supposed to be in tension with the principle of tolerance.[6] In light of this Friedman recommends that Carnap should restrict himself to weak metalanguages that are acceptable to all participants in the debate over mathematics:

5 I have changed the names, Carnap uses the less memorable 'X_1' and 'X_2'.

6 It is actually not so clear whether this example is really analogous to the case of Less and More: since classical logic can be interpreted in intuitionistic logic, it is doubtful whether classical logic is really unambiguously stronger than intuitionistic logic (Gödel 1986).

> By contrast, the choice of a restricted meta-language equally acceptable to all parties to the dispute is much better suited to Carnap's profession of tolerance. [...] We thus can see, even in this restricted metaframework, that the classical rules are much more expedient for physical applications, whereas the intuitionistic rules provide far more safety against contradiction. Hence, in accordance with the spirit of the principle of tolerance, we then can view the choice between the two frameworks as a fundamentally pragmatic one. (Friedman 1999b: 230)

Friedman adds that this conception undermines Carnap's way of drawing the analytic/synthetic distinction, but seems to think that preserving the spirit of tolerance is more important. But why does Friedman think that accepting a metalanguage in which classical mathematics can shown to be analytic rules out intuitionism in the first place? The only reason I can think of is an argument along the following lines: in such a metalanguage we can show that bivalence holds for mathematical discourse, and therefore intuitionism is immediately refuted. I think that this is much too quick, however, since there is a Carnapian response to the alleged incompatibility of strong metalanguages and the principle of tolerance.

Concerning the case of Less and More from above, Carnap comments as follows:

> I would object only if [Less] were to say to [More]: "In contrast to you, there is no possibility for me to choose between the two languages. On the basis of careful considerations, I have arrived at the following two ontological results:
>
> (6) There are classes of objects.
>
> (7) There are no classes of classes of objects.
>
> What you regard as semantical rules for L_1 contains the phrase 'classes of classes of objects', which does not refer to anything. Therefore, no semantical rules for L_1 have actually been stated; thus L_1 is not an interpreted language but merely a calculus". (Carnap 1963b: 873)

Note that the issue here is not that (6) and (7) are *false* – Less's language is in fact designed in such a way that they come out true. The problem rather arises if one wants to use (6) and (7) to argue *against* the language More uses. This attempt clashes with the distinction between external and internal questions, for while (6) and (7) are true in *Less's* framework, it cannot be a general requirement on accepting *any* framework that (6) and (7) come out true. If Less felt like it, they could adopt More's framework without making any factual mistake, which is just a consequence of Carnap's assumption that all external questions are pragmatic.

How is this relevant to the case of classical mathematics versus intuitionism? It seems to me that Carnap would reject Friedman's claim that the fact that in a classical metalanguage bivalence can be established is a refutation of intutionism. For it is perfectly possible to adopt an intuitionistic metalanguage as well, in which bivalence doesn't hold. And a classical mathematician who accepts the principle of tolerance and the internal/external distinction would not be in a position to call the adoption of an intuitionist metalanguage *incorrect* in any objective sense – once again, choosing between the two options is just a matter of pragmatic decision.

I therefore think that the second version of Friedman's argument fails as well, since the conflict he diagnoses is merely apparent. On the other hand, the thought that the "spirit of the principle of tolerance" is better served by the adoption of weak metalanguages is certainly not completely misguided. For as Carnap understands it, the principle of tolerance does seem biased in favour of strong frameworks, unless we have positive reason to suspect that the strong frameworks are inconsistent. This is because anyone who sticks with a weak linguistic framework, while not making a factual mistake, will quickly appear stubborn and irrational, given that Carnap does not believe in any metaphysical reasons for rejecting the conveniences of strong metalanguages. Goldfarb and Ricketts also flag this bias towards classical mathematics in Carnap's thinking:

> It should be noted, moreover, that from Carnap's vantage point only obscurantism prevents any logician from embracing the Principle of Tolerance. A proponent of intuitionistic mathematics who acquires an attitude of tolerance will not scruple at using classical syntax languages for matters that cannot be treated in weaker syntax languages. (Of course, the Principle of Tolerance does show Carnap's antipathy towards intuitionist criticisms of classical mathematics.) (Goldfarb and Ricketts 1992: 69)

It is therefore not unreasonable to wonder whether one can understand the principle of tolerance in a more neutral way, and anyone interested in this idea will find Friedman's appeal for weak metalanguages attractive. But as an objection to Carnap's actual position it has little force.

More recently Friedman has accepted that the second argument ”misses the essence of Carnap's conception” (Friedman 2009: 241). Not only that, but he now also holds that Beth's own argument, which motivated Friedman's initial skepticism about the use of strong metalanguages, is misguided. His response to Beth is now similar to that of Carnap, namely to point out that there is nothing mysterious or philosophically problematic about the need for a standard interpretation:

> [...] Carnap is completely untroubled by [... Beth's argument] because he is assuming, entirely reasonably, that an unproblematic understanding of the standard model of arithmetic is encapsulated in ordinary mathematical usage. There is no deep mystery here – there is no need to puzzle ourselves over the question how we somehow force an uninterpreted formal calculus to designate or refer to an intended model. (Friedman 2009: 240)

I think that Friedman has moved into the wrong direction here, however. As I have argued in the previous chapter, it only seems possible to claim that such an understanding of the standard model is part of our ordinary use of mathematical language if we use a language with non-recursive and infinitary rules, and hence Carnap runs into the problem of insufficient dispositions. This leads us to another defence of Carnap, which I will discuss in the next section: according to Gary Ebbs, there is a sense in which claims about metalanguages are *independent* of claims about dispositions, contrary to what I have been supposing.

5.2.2 Who Is Speaking?

My reconstruction of Beth's argument from non-standard models relies on the assumption that Carnap thinks that there is a fact of the matter concerning which metalanguage we actually speak, namely that it is like *ML* and unlike *ML**. This leads to the problem that while there is a sense in which recursive rules can correspond to linguistic dispositions, this kind of correspondence does not easily transfer to infinitary rules. I will now consider a response to this argument based on work by Gary Ebbs, who suggests that Carnap *can* say that we actually speak *ML without* needing to back up this claim by relying on the presence of particular linguistic dispositions.

This proposal is an interesting one, in particular since among all commentators Ebbs's interpretation of Beth's argument comes closest to the one I have presented – the main difference being that the distinction between recursive and non-recursive rules plays no role. Like me Ebbs notes that the undetermination of frameworks by linguistic disposition seems to lead to the potentially disastrous result that it is indeterminate whether speakers ever share a particular metalanguage or not:

> The trouble is that, by Carnap's own standards, a speaker's linguistic behavior does not *uniquely* determine what semantical rules she is following: there is more than one coherent set of coordinative definitions, hence more than one coherent semantical description of her linguistic behavior. This apparently undermines Carnap's starting assumption that investigators can share a metalanguage within which to codify semantical rules. (Ebbs 1997: 124f)

Unlike me, however, Ebbs thinks that this problem is merely apparent, since it is based on the failure to appreciate a crucial distinction Carnap makes: namely the distinction between *pure* and *descriptive* semantics. It is introduced as follows:

> By **descriptive semantics** we mean the description and analysis of the semantical features either of some particular historically given language, e. g. French, or all historically given languages in general. [...] On the other hand, we may set up a system of semantical rules, whether in close connection with a historically given language or freely invented; we call this a *semantical system*. The construction and analysis of semantical systems is called **pure semantics**. (Carnap 1942: 11f)

Given this terminology, the task of coordinating a particular linguistic framework with an actually spoken language is obviously part of descriptive semantics. The construction of some arbitrary linguistic framework without any regard for whether it corresponds to some language already in use, on the other hand, would presumably be an example of pure semantics. In order to flesh out the distinction some more, Carnap draws an analogy to geometry:

> Both in semantics and in syntax the relation between the descriptive and the pure field is perfectly analogous to the relation between pure and mathematical geometry, which is a part of mathematics and

> hence analytic, and physical geometry, which is a part of physics and hence empirical (Carnap 1942: 12)

I take the analogy Carnap has in mind here to be as follows: we can construct various kinds of geometrical calculi without having to care about the geometry of *physical space*. If we do so we either treat the geometrical systems as completely uninterpreted, or as having a purely mathematical interpretation. That's pure geometry, and it would be inappropriate to object to or argue in favour of any system of geometry understood in this sense. Similarly in pure semantics: maybe we have constructed a linguistic framework that doesn't correspond to any actually used language, but if that was never the aim there is nothing wrong with it. The descriptive case is different: if we are interested in capturing the structure of physical space, then obviously not all systems of geometry are equally suited to this purpose. Empirical facts will rule out many geometries, and equally so in the case of descriptive semantics: not just any arbitrary linguistic framework can be coordinated with a natural language like English.

Using this new terminology, we can observe that I have construed Carnap's *we in fact speak ML* response as a thesis of descriptive semantics, and hence empirical facts about what kinds of dispositions we do and can have were relevant. According to Ebbs, however, this is actually a mistake:

> This "problem" stems from a subtle but serious misunderstanding of both pure and descriptive semantics. For Carnap an empirical description of the semantical properties of a speaker's utterances is not a description of how the speaker implicitly "interprets" her words. Empirical discoveries about the semantical properties of a speaker's utterances can neither *undermine* nor *justify* agreements reached within pure semantics, since such agreements set the ultimate parameters for our inquiries. In this sense there is no legitimate perspective higher or firmer than pure semantics from which to question our use of the shared metalanguages to codify rules for inquiry. (Ebbs 1997: 125)

Unpacking Ebbs's understanding of Carnap's position here is not easy. One might think that he wants to argue as follows: the claim that we use and share metalanguage *ML* is not a claim of descriptive semantics, but rather of pure semantics. For this reason empirical facts about actual human beings are irrelevant and can neither justify nor undermine it. And thus the considerations about dispositions I presented earlier are unsuited to put pressure on Carnap's claim that we actually speak *ML*. Hence Carnap's reply to Beth remains unscathed after all.

The problem with this response, however, is that it is hard to understand how the claim that we speak metalanguage *ML* could possibly be a claim of pure semantics. For it seems to essentially involve reference to a community of speakers (the *we*), and, so I assume, these speakers are human beings, and hence part of the empirical world. Pure semantics was supposed to be independent of facts about the empirical world, however, and therefore cannot really involve reference to speakers at all, just as pure geometry is not a theory about physical space. If this is correct, it would therefore not even be possible to express the claim that we speak a particular metalanguage within pure semantics.

In personal communication Ebbs has clarified that he actually had a different kind of reply in mind. The thought is not that claims about metalanguages are part of pure semantic, but rather that we need to *presuppose* that we speak and share a certain metalanguage before we can even begin to do any semantic investigations, whether pure or descriptive. There is thus a sense in which the actual use of a metalanguage can neither be justified nor undermined, since we can only reason about our metalanguage from the perspective of this very metalanguage. Ebbs therefore describes our actual use of a metalanguage as a *pragmatic presupposition*, and draws a parallel to Quine's acquiescence in our home language.

To a point I agree with Ebbs, for it would surely be unreasonable to question our sharing of a metalanguage on the basis that there is no conclusive proof establishing this fact. As long as communication proceeds smoothly it is indeed sensible to assume that we speak the same language. But this kind of consideration only establishes that we are entitled to the pragmatic presupposition that we speak and share *some* metalanguage or other. As I stressed in section 4.2.1, however, enabling communication is not the only role Carnap has for a metalanguage: it is also supposed to resolve Gödelian incompleteness, and it is this latter feature that Beth attacks. As far as I can see, there is no way to run Ebbs' argument in such a way that it entitles us to assume that our metalanguage is as strong as *ML* or *ML**. For it is arguably not the case that we can only communicate if we use a metalanguage with non-recursive inference rules, rather than one with only recursive rules that would not allow us to overcome incompleteness. I therefore think that the claim that we speak *ML* goes beyond what can reasonably be taken for granted without further argument, and hence questions about linguistic dispositions remain salient.[7]

[7] Comparing Carnap to Quine is illuminating in this context, since I think that the latter can indeed rebut any Beth-style worries in the way Ebbs suggests. This is possible precisely because Quine does not want the metalanguage to settle the truth values of mathematical

For this reason I think that Ebbs' defence of Carnap does not suffice to undermine Beth's argument. I will conclude this section by briefly considering another way of cashing out the idea that facts about metalanguages are independent of facts about linguistic dispositions, namely by construing the claim that we speak metalanguage *ML* in such a way that it does not refer to empirical human beings at all. This may seem outlandish at first, but there is a long philosophical tradition of distinguishing between an empirical and some kind of non-empirical self – Kant's distinction between the empirical and the transcendental self is probably the best-known example. And this idea can also be found among Carnap's contemporaries, for instance in Wittgenstein's *Tractatus*:

> Thus there really is a sense in which philosophy can talk about the self in a non-psychological way.
>
> What brings the self into philosophy is the fact that 'the world is my world'.
>
> The philosophical self is not the human being, not the human body, or the human soul, with which psychology deals, but rather the metaphysical subject, the limit of the world – not a part of it. (Wittgenstein 1961: 5.641)

One could therefore interpret Ebbs as follows: when Carnap says that we speak metalanguage *ML*, the 'we' in this claim doesn't refer to empirical human beings at all, but to speakers of a language in a non-empirical sense. And for this reason concerns about dispositions are misplaced.

A lot could be said about this suggestion, but I think it is pretty clear that Carnap would not have openly endorsed anything like non-empirical selves. For while he was strongly influenced by the *Tractatus*, he quickly came to reject the more esoteric and metaphysical aspects of the book. As Carnap describes in his Intellectual Autobiography, this difference in view especially applies to Wittgenstein's conception of language:

> We read in Wittgenstein's book that certain things show themselves but cannot be said; for example the logical structure of sentences and the relation between the language and the world. In opposition to this view, first tentatively, then more and more clearly, our conception developed that it is possible to talk meaningfully about language and

statements, and so for him enabling communication is the only relevant task. I will come back to the relationship between Beth and Quine in chapter 8.

> about the relation between a sentence and the fact described. Neurath emphasized from the beginning that language phenomena are events *within* the world, not something that refers to the world from outside. (Carnap 1963a: 29)

I think that this attitude is incompatible with a notion of language that is independent of what happens in the empirical world, and hence incompatible with non-empirical selves.

In the end I therefore see no way to use the distinction between descriptive and pure semantics in such a way that Carnap can maintain his reply to Beth without facing the problems caused by the lack of dispositions for non-recursive rules. One could of course conclude that Carnap was mistaken to reject non-empirical selves à la Wittgenstein, and should have endorsed the view I just considered. This may well be right, but then this proposal would not be so much a *defence* of Carnap's actual position, but rather a *rejection* of a central tenet of his philosophy: namely his strictly anti-metaphysical stance. For now I therefore conclude that Beth's challenge works as it stands.

5.2.3 Explication

The argument I presented by drawing on Beth is basically an *underdetermination argument*: different definitions of analyticity for mathematics which are not extensionally equivalent are all equally compatible with our linguistic dispositions, and hence no one definition of analytic is pinned down by the latter. André Carus has recently argued that focusing on the underdetermination of formal languages by linguistic dispositions is misguided, however, since it prevents us from appreciating that Carnap's real aim is to *reform* language rather than just to *describe* it. This is supposed to be a consequence of Carnap's idea of *explication*, and applied to the case at hand the thought is that Carnap's account of analyticity is not a descriptive claim about actual language use at all, but rather as a *proposal* about how we *should* speak:

> Proposed Analyticity
> For all mathematical sentences S: *we should talk in such a way* that S is either analytic or contradictory.

In the end I will argue that this defence is unsuccessful, since the lack of dispositions for non-recursive rules also undermines Proposed Analyticity. But since the notion of an explication plays a major role in Carnap's later philosophy it will be worth going through the proposal in some detail, as it will further clarify the theoretical role linguistic frameworks are supposed to play.

Let us us begin by looking at Carnap's own explanation of what it means to explicate a concept:

> The task of *explication* consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. We call the given concept (or the term used for it) the *explicandum*, and the exact concept proposed to take the place of the first (or the term proposed for it) the *explicatum*. The explicandum may belong to everyday language or to a previous stage in the development of scientific language. The explicatum must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific either logicomathematical or empirical concepts. (Carnap 1950: 3)

He then stresses that since the target concept of natural language, i.e. the explicandum, will be inexact, it does not make sense to ask whether a proposed explicatum is *correct* or not. Carnap does give some criteria for what a *good* explication should strive for though:

> 1. The explicatum is to be *similar to the explicandum* in such a way that, in most cases in which the explicandum has been so far used, the explicatum can be used; however, close similarity is not required and considerable differences are permitted.
> 2. The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an *exact* form, so as to introduce the explicatum into a well-connected system of scientific concepts.
> 3. The explicatum is to be a *fruitful* concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept).
> 4. The explicatum should be as *simple* as possible; this means as simple as the more important requirements (1), (2), and (3) permit. (Carnap 1950: 7, §3)

(1) is of particular interest to us, since Carnap allows deviations from ordinary usage. The following reply to Beth's argument suggests itself: Carnap's formal

definition of *analytic* for mathematics is clearly intended to be an explication of the vague natural language concept. The problem is that our actual usage of the word does not pin down the definition Carnap suggests over extensionally deviant explications. But it is clear that Carnap did not regard this as a requirement for good explications anyway, and so the disposition argument is irrelevant given Carnap's own conception.

Carus' discussion supports the kind of reply I just summarised. He argues that one could of course construct formal systems of syntax and semantics, including a definition of analyticity, with the aim of capturing and predicting existing ordinary usage. But, so he stresses, this project is quite different from giving explications, and hence objections to the one kind of project do necessarily transfer to the other:

> Quite separately from (and irrelevantly to) this, however, pure semantic or syntactic theories (with or without empirical models) may be put forward as *explications* of vague concepts in ordinary language, such as the logical words that indicate connections among segments of everyday speech that could by some behavioural standard be classed as 'deductive'. An empirically interpreted semantic theory may succeed or fail as an empirical hypothesis, but this has no bearing on its corresponding purely logical theory as a candidate for *explicating* such vague concepts of ordinary language. (Carus 2007: 248)

In order to assess whether this is a successful rebuttal of Beth, we need to have a better understanding of what the *point* of explications is. Suppose we have given an explication of some informal everyday concept by describing explicit rules of use for the replacement concept in a formal system. What happens next? Presumably the idea is that we can *use* the explicated concept, either by changing the way we talk in everyday life, or at least by using it in scientific contexts where more precision than in ordinary language is desirable. According to Carus, explications give rise to an interplay between natural language and formal systems that represents a major paradigm change in Carnap's overall thinking. The project of rational reconstruction Carnap pursues in the *Aufbau*, for instance, is more unidirectional in that it always privileges scientific notions over those of ordinary language:

> Unlike rational reconstruction, explication no longer envisaged one-way replacement of the ordinary, intuitive world view by a scientific

> one, but a dialectical interchange between the two kinds of system. Our practices and our values reside within an intuitive *Lebenswelt* that can be progressively improved, whose quality can be raised piecemeal through explicative replacement of its concepts by constructed ones, but we decide what replacements to undertake from the overall standpoint of the *Lebenswelt*, our practical concerns and our values. (Carus 2007: xi)

While the overall picture of Carnap's project of explication presented here sounds straightforward and compelling, the idea that we can use and adopt certain explications and thereby modify natural language needs further attention. One of the examples of explications Carnap himself gives concerns the concept *temperature*: he describes the development of quantitative temperature concepts from the more basic concepts found in ordinary language, such as 'hot' or 'warmer than' (Carnap 1950: 9f). This is helpful, for it is not hard to see what it would mean for a population of speakers who do not already have quantitative temperature concepts to adopt them into their everyday life. We thus have a case in which something that begins as a purely formal description of a certain concepts eventually influences the way people talk.

How about Carnap's explication of analyticity, however? We already saw that his preferred explication determines each mathematical sentence to be either analytic or contradictory, which means that it settles the status of all the infinitely many sentences not decided by the Peano axioms. We also saw that it is implausible to maintain that we *actually have* dispositions to use 'analytic' in this way. This didn't seem so bad in light of the fact that explications don't need to conform to prior usage. Suppose now, however, that we like Carnap's explication and want to adopt it. This requires that we *change* our dispositions in such a way as to conform to the explication. But this is arguably impossible, for the same reason that claiming that we already *have* these dispositions is not credible: namely that it in general seems impossible for finite human beings to have dispositions which are non-recursive.

It is very unfortunate that, despite Carnap's heavy reliance on non-recursive rules in his philosophy of mathematics, he rarely problematises their philosophical implications. There are reports that he did so when teaching, but little is known about the content of these discussions:

> In classroom discussion, when pursued by questions as to how the 'ultimate' metalanguage came to be understood, Carnap would ex-

> pand at length on how languages carne to be learned by children – first by pointings, then by context, correctings, and by increasing use of informal metalinguistic talk (in a manner not unlike Quinean regimentation). On a few occasions, this was extended into a revealing discussion of how Language II, with its apparently non-intuitive number-based syntax, would be learned by a child in a society which spoke such a language. (Bohnert 1975: 196)

It might well be that Carnap had some account of how to use infinitary rules that he never discussed in his writings. But to me it seems more likely that he just did not think that much about what he calls *pragmatics*, i.e. the study of spoken languages and their relation to formal systems, since he preferred the construction of formal systems. There is at least one other case where a critic of Carnap – in this instance Otto Neurath – was dissatisfied with a language Carnap proposed on account of it being unusable in practice. As is well-known, in the *Aufbau* Carnap tries to construct the physical world in an *auto-psychological language*, i.e. one in which I ultimately only talk about my own experiences (Carnap 1967). In the 1930s Carnap then moved towards preferring a physical language as the base language, and even gave something like a private language argument (Carnap 1934). Neurath was a driving force in this development, for he had advocated using a physical language early on (Neurath 1931).

For Neurath, however, Carnap's rejection of auto-psychological languages did not go far enough. Carnap's private language argument is only meant to establish that each sentence of a phenomenal language has the same truth-conditions as a statement in a physical language. And while Carnap now expresses a preference for physical languages, he still considers the choice to be a pragmatic matter. Neurath, on the other hand, wants to say something stronger: for him it is misleading to call the phenomenal language of the *Aufbau* a genuine language at all, since it is not something we could actually speak. Thomas Uebel sums this up concisely:

> Neurath did not argue that phenomenal languages were *logically impossible*: Carnap had apparently shown how to construct one in the *Aufbau* (the technicalities of which Neurath did not fault). [...] Any protocol languages whose sentences needed no justification were to be rejected as *"non-realizable."* This formulation clearly takes account of, and rejects as irrelevant, Carnap's proof of the logical possibility of an auto-psychological language, namely its construction in the ab-

> stract. A private language was – in a sense to be made precise – *a human impossibility*. (Uebel 1992: 453)

We do not need to go into how Neurath spells out the relevant possibility here any further, but I think that the general lesson Uebel draws from this disagreement is correct: unlike Neurath, Carnap is happy to describe systems of rules as languages even though it is not humanly possible to follow their rules. And while this may be fine in some contexts, it becomes a problem when we want to *adopt* such a system of rules as an explication. If explications are not just given for the sake of doing so, but need to be usable by us, then Carnap's explication of analytic for mathematics fails this crucial test.

My conclusion is therefore that, even if Carus is right that using a formal system as an explication is very different from using it as a description of actual behaviour, the problem of insufficient dispositions nevertheless remains. Reading Carnap's definition as a *proposal* rather than a descriptive claim may help against some objections that have been raised, but not against the version of Beth's argument I have developed.

Chapter 6

Radical Deflationism

In this chapter I will discuss a more radical attempt to defend Carnap's position against Beth's argument: namely rejecting the fundamental assumption that, as a matter of fact, we actually use metalanguage *ML*. This kind of response is inspired by the deflationary readings of Carnap we already encountered in section 1.2.3. I will argue, however, that such an interpretation of Carnap is not only exegetically implausible, but also results in a much more extreme position than deflationary commentators such as Goldfarb and Ricketts recommend. This will be illustrated by comparing the resulting philosophy of mathematics to Wittgenstein's controversial radical conventionalism.

6.1 Carnap and Semantic Facts

My reconstruction of Beth's argument can be summed up as follows: for Carnap to maintain that we in fact speak metalanguage *ML*, which is committed to the standard model of arithmetic, he needs to hold that our linguistic dispositions encode a commitment to the standard model. But, given plausible assumptions, our linguistic dispositions must be describable using a *recursive theory*, and since by Gödel's incompleteness theorems all such theories have non-standard models, the task cannot be achieved.

This reading presupposes that Carnap himself actually accepts the existence of certain *semantic facts*, namely those of the form

> A population of speakers (such as ourselves) speaks one particular metalanguage *ML* rather than another.

This assumption needs to be scrutinised more carefully, however. For in effect the crucial question is whether our speech behaviour pins down a particular linguistic framework, and in section 4.1.3 I already flagged that Carnap's own remarks on this issue are not clear-cut.

Moreover, a version of the question whether Carnap is committed to semantic facts has already been controversially discussed in the secondary literature. In his influential book on *The Semantic Tradition from Kant to Carnap* (Coffa 1991), J. Albert Coffa asks whether, for Carnap, it is a fact that mathematics is merely a matter of conventions. Coffa argues for an affirmative answer, and ascribes to Carnap a *factualism at the second level*, according to which there are semantic facts about what is conventional or not:

> The multiplicative axiom is not a factual claim but a convention. But *this* statement is not a proposal for a convention. It is a factual statement about the nature of mathematical axioms. (Coffa 1991: 322)

Goldfarb has convincingly argued, however, that it is implausible to interpret Carnap in this way. For Coffa's reasoning seems to be as follows: if there were *no* semantic facts of the relevant kind, there would be no sense in which Carnap's preferred account of mathematics is *more correct* than that of a mathematical realist, such as the external Platonist we have frequently considered. Against this, Goldfarb stresses the *deflationary nature* of Carnap's approach to philosophical questions, on which factual questions about which view is objectively correct are replaced with comparisons based on pragmatic considerations:

> I am suggesting that Carnap's position in [... *Logical Syntax*] is *deflationary*. [...] Its stance toward alternative philosophical positions is not of opposition to their doctrines with its doctrines, but of invitation to clarification. The result of those clarifications, the frameworks of the alternative positions, can then be compared with Carnap's favored frameworks, and the greater desirability of Carnap's, presumably, will be evident. It is important to realize that those comparisons can be made only on a case-by-case basis: there is no general theory invoked against the alternatives. All of Carnap's considerations against other philosophical positions must be ad hoc. Indeed, no other method would be compatible with thoroughgoing application of the Principle of Tolerance. (Goldfarb 1997: 61)

As I understand this, Goldfarb's deflationary reading amounts to the following: when Carnap claims that mathematical axioms are conventional, this is not supposed to entail that anyone who adopts a linguistic framework in which mathematics is *not* analytic is making some kind of factual mistake. Carnap rather *recommends* that we adopt linguistic frameworks in which mathematics is analytic, in the hope that we come to appreciate the convenience of this move. Contra Coffa, there is thus no exception to the claim that choosing frameworks is a matter of pragmatic decision.

This strikes me as a compelling reply to Coffa. The question that arises, however, is whether a similar response could be given to Beth. In the following I will argue for a negative answer: to avoid the problem of non-recursive dispositions, a much more radical form of deflationism would be required. In the following I will explain why this is the case, and then argue that radical deflationism is exegetically implausible, before considering the view on its own merits in the next section.

In the quote above, Goldfarb describes Carnap's strategy as a comparison between different frameworks, with Carnap hoping that his own preferred framework for mathematics will prove to be most attractive. For this description to make sense, however, it seems that one needs to be in a position to adopt and use Carnap's preferred framework. For if not, it is unclear what the point of constructing the framework was in the first place. And this observation brings us right back to Beth's objection. For the problem with Carnap's employment of non-recursive inference rules in his framework for mathematics is precisely that it is unclear how human beings can actually *use* such a framework. Goldfarb's response to Coffa thus *presupposes* that there can be semantic facts such as "we speak metalanguage *ML* rather than another", and therefore cannot answer any worries about the possibility of such facts.[1]

This is not to say that there is *no* way to give a deflationary response to Beth's argument. There is indeed an obvious candidate: one could deny that Carnap really wants to claim that we determinately speak metalanguage *ML* rather than *ML**. In other words, we could read Carnap as agreeing that our recursive dispositions don't uniquely pin down a certain framework with non-recursive rules, and that hence which framework we speak is to a certain extent indeterminate.

On an exegetical level this proposal might seem to be a non-starter, for when discussing Carnap's reply to Beth from the Schilpp-volume, I emphasised that

[1] One might worry that my reliance on semantic facts is in tension with Goldfarb's rejection of a language-transcendent notion of fact, which we encountered in his (and Ricketts') response to Gödel (section 2.2.2). I think that this is not the case, however. The relevant semantic facts are rather internal to the framework of *descriptive semantics*, in which we can discuss the coordination of dispositions with linguistic rules.

his main point is that we in fact speak one particular metalanguage, namely *ML*, rather than *ML**. It must be admitted, however, that the reply also contains passages which point in a different direction. Carnap there explicitly addresses Beth's contention that phrases such as "for all syntactical properties of accented expressions" and "and so on", which occur at crucial junctures in *Logical Syntax*, can be interpreted in different ways:

> It is of course not quite possible to use ordinary language with a perfectly fixed interpretation, because of the inevitable vagueness and ambiguity of ordinary words. Nevertheless it is at least possible to approximate a fixed interpretation to a certain extent, e.g., by a suitable choice of less vague words and by suitable paraphrases. [...] With regard to this difficulty Beth gives two good examples. First he refers to a passage in [Syntax] (p. 13) which contains in an informal explanation the phrase "and so on" [...], and secondly he refers to a place (p. 113) where I myself point out that phrases like "for all syntactical properties of accented expressions", which occur in syntactical rules, are ambiguous. (Carnap 1963b: 930)

This case is crucial to the issue of non-standard interpretations, since the relevant occurrence of "and so on" appears in a passage where Carnap enumerates the numerals $0, 0', 0''$, ..., (Carnap 1937a: 13). It is thus ambiguous insofar as one could continue this enumeration in standard *or* non-standard ways. One might now think that, if Carnap actually endorsed the claim that our metalanguage *ML* encodes a commitment to the standard numbers, he should deny that the relevant ambiguity exists. In fact, however, he appears to be in complete agreement with Beth:

> I would certainly agree if Beth had said about these examples something like this: "At these two places the requirement of strict usage of natural language, i.e., of a usage of words with fixed meanings, is not fulfilled". Amazingly, he says just the opposite; [...] Later he says: "The term 'and so on' which appears in Carnap's text, is supposed to be univocal". Presumably he means hereby to imply: "but, in fact, it is not univocal"; and with this I would agree. (Carnap 1963b: 930f)

It is certainly possible to interpret this passage as saying that our ordinary use of mathematical language does not pin down the standard numerals to the exclusion of non-standard numerals. And if this is the correct reading, then the conclusion that, according to Carnap, there is no fact of the matter whether we speak metalanguage *ML* or a non-standard version such as *ML** is indeed very natural. Overall I think that the reasons against such an interpretation outweigh the reasons in favour, however. Note that holding that, in the relevant context, "and so on" is ambiguous between standard and non-standard readings is compatible with believing that we manage to *resolve* this ambiguity in a particular way. I therefore think that what Carnap agrees to here is that it is *possible* to interpret the statements found in *Logical Syntax* in a non-standard way, but not the stronger claim that it is *impossible* to determinately pin down the standard conception. This would speak against reading Carnap as a radical deflationist, and I will support my interpretation by drawing on an as yet unpublished letter from Carnap to Irving Copi on the philosophical implication of Gödel's incompleteness results.

In 1950, Copi published an article called "Modern Logic and the Synthetic A Priori" (Copi 1949). Copi there argues that, in light of Gödel's incompleteness theorems, the popular view that all mathematical truths are analytic must be given up, at least if analyticity is identified with truth in virtue of linguistic rules. He then suggests that undecidable sentences should be considered *synthetic* a priori truths. This paper spawned a short debate in the *Journal of Philosophy* where Copi's article appeared. In a response, Atwelle Turquette made the alternative suggestion that we should not regard undecidable mathematical sentences as determinately true or false at all, but that calling them *indeterminate* is more appropriate – a contention Turquette supports by considering the possibility of non-standard models, referring to Henkin's dissertation (Turquette 1950). In fact Turquette goes even further and suggests that sentences such as Con_{PA} are *meaningless* and *nonsensical*, a proposal Copi rejects as implausible in a further reply (Copi 1950).

Carnap did not publicly intervene in this debate, but he apparently regarded it as sufficiently important to write Copi a letter in which he defends his own view that all mathematical truths are analytic. (Unfortunately for our purposes, Carnap doesn't directly comment on Turquette's paper). In a move that will not surprise us, the gist of Carnap's letter is that Copi's conclusion that undecidable sentences do not follow from linguistic rules relies on a too narrow construal of what kind of rules are relevant. The following is a central passage:

> [The concept "analytic"] is by no means identical with "provable", because it is based on transfinite syntactical rules. [...] The decisive point which you seem to overlook is this: it follows from Gödel's

> consideration that his undecidable sentence is true and, moreover, L-true. In other words, Gödel has actually shown, on the basis of linguistic rules, (though not on the basis of the syntactical rules of the given calculus), that the sentence is true. Thus the sentence is known to be analytic. (Carnap, Letter to Copi from August 23 1949, RCP 088-18-08)[2]

I think that this remark strongly supports a reading of Carnap on which he *does* think that we in fact speak metalanguage *ML*, for a number of reasons. First, the linguistic rules he is alluding to here are clearly those of a metalanguage. It is therefore certain that he holds that we speak a metalanguage that settles the truth values of *some* undecidable sentences. Secondly, I think it is reasonable to assume that the particular Gödel sentence Carnap is talking about here is just a convenient example, for nowhere in his work does he draw a theoretically important distinction between different kinds of undecidable sentences. If this is so, then the point he makes generalises: the linguistic rules of our metalanguage do not only settle the truth value of one, but of *all* undecidable sentences. Thirdly, it is striking that the letter is formulated in a way that suggests Copi has made a *factual* mistake: Carnap does not merely say that we *could* accept linguistic rules that make the Gödel sentence come out as analytic, but he rather implies that we already *are* committed to such rules anyway.[3]

The reading of Carnap's position on the analyticity of mathematics that emerges from the letter to Copi, and also large parts of his reply to Beth, can thus be captured as follows:

> Absolute Analyticity
> Given the linguistic framework we actually accept, the following holds: for all mathematical sentences *S*, it is determinate that *S* is either analytic or contradictory.

This position requires that it is determinate which framework we accept, which in turn relies on facts about what metalanguage we speak.

2 The Rudolf Carnap Papers are based at the Archives of Scientific Philosophy of the University of Pittsburgh.

3 This is worth stressing since in many instances Carnap is keen to reinterpret apparent factual disputes about philosophical matters as different proposals for how to speak. He does so explicitly even for issues like monism versus dualism in the philosophy of mind (Carnap 1963b: 884f).

6.2 Wittgenstein's Radical Conventionalism

Earlier I described a deflationary response that denies facts about what metalanguage we use as *radical*, and we are now in a better position to appreciate why. Without the relevant facts, Carnap's position on the analyticity of mathematics amounts to the following:

> Relative Analyticity
> For some mathematical sentence S, the linguistic framework we actually accept does not determinately settle whether S is analytic or contradictory. Moreover, S is analytic relative to some way of extending our linguistic framework, and contradictory relative to another.

And this seems to be quite a major departure from Carnap's official position. For there will now be purely mathematical sentences that are indeterminate given the framework we actually accept, a status that Carnap had originally reserved for *synthetic* sentences about the empirical world (Carnap 1937a: 28). It might of course be that this revision of Carnap's view is nevertheless the best way of defending him against Beth, and in this section I will therefore discuss what further costs this move has.

I characterised radical deflationism as the view that which linguistic framework for mathematics we accept is to a certain extent indeterminate. In order to assess how viable this position is, we need to consider the question of how far this indeterminacy reaches. Given the argument from dispositions of the previous chapter, the only defensible option seems to be the following: the recursive rules of the framework of mathematics are settled by our usage of mathematical language, but the non-recursive rules are left indeterminate.

In their article on Carnap's philosophy of mathematics, Goldfarb and Ricketts discuss what would happen to Carnap's position if he could only use *weak* metalanguages, i.e. ones with recursive rules. This is obviously a version of the view that only the recursive rules of a framework are determinate, and so the conclusion they draw is immediately relevant for the proposal at hand:

> This way of characterizing how little is left of the notion of mathematical truth shows how far from Carnap's stated views the restriction to weak metalanguages winds up being. It summons a rather different philosopher to mind. (Goldfarb and Ricketts 1992: 77)

The philosopher they allude to is Wittgenstein during his so-called middle period, and the views he explores in the *Remarks on the Foundations of Mathematics* and lectures at Cambridge. They quote the following striking passage, in which Wittgenstein sums up his interpretation of the philosophical upshot of the incompleteness theorems:

> "But may there not be true propositions which are written in this symbolism, but are not provable in Russell's system" – 'True propositions,' hence propositions which are true in *another* system, i.e., can rightly be asserted in another game. Certainly; why should there not be such propositions [...] [... A] proposition which cannot be proved in Russell's system is "true" or "false" in a different sense from a proposition of *Principia Mathematica*. (Wittgenstein 1956: 50e)

The natural interpretation of these remarks is as follows: Wittgenstein denies that there is a unified notion of *mathematical truth*, if this is understood to go beyond what is derivable in some specific calculus. Instead, all he accepts is *truth-according-to Russell's system*, *truth-according-to-another-system*, etc.

I think that, when read as a radical deflationist, Carnap would have to say something very similar. If our dispositions don't pin down one particular framework for doing mathematics, but are rather compatible with a number of them, then it becomes hard to see how he can speak about *the* framework of mathematics at all. And while this result might be fine for some areas of mathematics, the consequences when applied to arithmetic are quite severe. For as we saw in section 4.2.3, without determinate truth values in arithmetic we do not get determinate facts about *syntax*, including facts concerning what is derivable from what.

Wittgenstein himself was aware of this peculiar consequence of his view, but he openly embraced it, and for this reason has been described as a *radical conventionalist*. The standard view among conventionalists about logic and mathematics is that once we have accepted certain conventions, it is thereby determined what does and doesn't follow from them. As Dummett points out, however, one might be unsatisfied with such accounts if the nature of this *determination* remains unexplained:

> This account is entirely superficial and throws away all the advantages of conventionalism, since it leaves unexplained the status of the assertion that certain conventions have certain consequences. [...] Wittgenstein goes in for a *full-blooded conventionalism*; for him the logical necessity of any statement is always the *direct* expression of a linguistic convention. (Dummett 1959: 328f)

As Dummett reads him, Wittgenstein holds the surprising view that the acceptance of certain axioms and inference rules do *not* compel us to accept statements which are derivable from them:

> [...] at each step we are free to choose to accept or reject the proof; there is nothing in our formulation of the axioms and of the rules of inference, and nothing in our minds when we accepted these before the proof was given, which of itself shows whether we shall accept the proof or not; and hence there is nothing which *forces* us to accept the proof. (Dummett 1959: 330)

Whether this is correct as an exegesis of Wittgenstein's philosophy of mathematics is controversial, and alternative interpretations have been offered.[4] There is no denying, however, that (at least in this period) Wittgenstein's views about the nature of syntax appear to be unlike those of pretty much any other logician and mathematician. It was and is widely accepted that for any system of axioms and inference rules, there is a fact of the matter whether they are consistent, i.e. whether a contradiction is derivable from them, regardless of whether anyone has ever *actually* derived a contradiction or not. This broad consensus is repeatedly challenged by Wittgenstein, however, for instance in conversation with Schlick and Waismann:

> One imagines that there could be a contradiction that no one has ever seen hidden in the axioms from the beginning, like tuberculosis. Suspecting nothing, one suddenly drops dead. And so it is thought, by analogy, that a hidden contradiction could erupt and bring catastrophe. (Waismann 1979: 120)

Wittgenstein takes up this health-based metaphor again in his lectures on the foundations of mathematics at Cambridge, where he rejects the view that finding a contradiction, "like finding a germ in an otherwise healthy body, shows that the whole system or body is diseased" (Wittgenstein 1989: 138). And his rejection of consistency facts that are independent of actual proofs is most explicit in the following passage:

> If you say, "The mere fact that a proof *could* be found is a fact about the mathematical world", you're comparing the mathematician to a man who has found out something about a realm of entities, the physics of

4 For discussion see Stroud 1965 and Putnam 1979.

> mathematical entities. If you say, "You can go this way or that way", you say there is no physics about mathematics. (Wittgenstein 1989: 138)

As ever when interpreting Wittgenstein one needs to be careful, for some of the radical-sounding claims found in his writing may not express his own position at all, but are rather vivid expressions of views Wittgenstein ultimately wants to reject.[5] Dummett was not the only one who took Wittgenstein to put forward strikingly unorthodox theses on the nature of mathematics, however. Gödel, for instance, is reported to have reacted to the publication of *Remarks on the Foundations of Mathematics* by asking whether Wittgenstein had lost his mind (Wang 1996: 179).[6] For our purposes I will put such exegetical questions aside, since it is instructive to consider radical conventionalism whether or not it was Wittgenstein's own view.

It should be uncontroversial that Carnap's position is not *actually* a form if radical conventionalism: in section 1.2.3 I already presented a quote in which Carnap characterises the truth of certain sentences as being "determined by the logical relations holding between the given meanings" (Carnap 1963b: 916), which seems to be just the kind of determination Wittgenstein rejects. And I think it is also clear that Carnap *should not* embrace radical conventionalism either, since it seems to undermine one of the most fundamental ideas behind the notion of a linguistic framework: namely that we can set up a framework by formulating explicit rules which then determine further sentences to be analytic, contradictory, or synthetic – hardly an incidental component of Carnap's position.

That Carnap needs there to be determinate facts about consistency can be demonstrated by looking back at Carnap's account of empirical content from section 2.2.2. We there encountered the following definition:

---

[5] Cora Diamond has for instance proposed a reading of Wittgenstein's philosophy of mathematics on which his position does not seem particularly revisionary (Diamond 1991).

[6] I am not sure whether Carnap was aware of Wittgenstein's thoughts on the philosophy of mathematics in the middle period. There is reason to doubt it, however, as the following anecdote illustrates:

> Years later some of *Wittgenstein's students at Cambridge* requested his permission to send transcripts of his *lectures* to friends and interested philosophers. He asked for the list of names, and then *approved all of them but mine*. In my life I have never experienced anything even similar to this hate against me. (Carnap, note dated 16.11.1956 entitled "This is *not* in the Autobiography!", RC 102-78-08)

> The content of a sentence *S* is the set of non-valid sentences which follows from *S*. (adapted from Carnap 1937a: 175 (§49)).

Whether there are any non-valid sentences or not, however, depends on whether the relevant linguistic framework is consistent. And without facts about consistency the whole approach would therefore be undermined, for it seems incoherent to assume that whether a framework has any empirical content at all could be either indeterminate or a matter of choice. For Carnap's position to be a stable one, the collapse into radical conventionalism thus needs to be resisted.

6.3 Is there a Middle Way?

I have so far suggested that the choice between Relative and Absolute Analyticity is exclusive. In this section I will consider a middle way between these two extremes, which has has recently been defended in a number of papers by Gregory Lavers. The general idea is that while for some areas of mathematics, such as set theory, it is very plausible that our linguistic dispositions do not commit us to one particular framework, an absolute conception might still be viable for arithmetic. Along the way I will also address Lavers' claim that Carnap's philosophy of mathematics underwent a major change between *Logical Syntax* and the 1950s.

Lavers is very clear that, when read unrestrictedly as applying to all of mathematics, he rejects what I have called the absolute conception of analyticity, according to which our linguistic practices commit us to a determinate way of dividing sentences into analytic and contradictory:

> We are able to show that for any mathematical sentence *S*, either *S* is analytic or *S* is contradictory. But this does not imply that our conventions assign a truth value to every sentence in some absolute sense: the admission that the English expressions in the metalanguage are vague rules out this possibility. (Lavers 2008: 22)

At times it sounds as if Lavers' reaction to this is simply to endorse the relative conception of analyticity, by maintaining that undecidable mathematical sentences have no determinate truth value. Commenting on Carnap's reply to Beth, Lavers suggests that Carnap does not actually want to use infinitary methods to secure determinate truth values for all of mathematics, but is rather content to live with a certain amount of indeterminacy:

> So long as the meaning of the English expression 'natural number' is determined by a finite set of rules, [...] our ability to use Language II is also, at bottom, explained by the grasp of a finite set of rules. Carnap's conventionalism can be saved by conceding that our mathematical notions are somewhat vague. (Lavers 2004: 313)

This initial impression is misleading, however, as Lavers' considered position is more sophisticated. An important notion in his overall interpretation of Carnaps' mature – i.e. 1950s – philosophy of mathematics is that of an *intuitive conception of mathematical truth*. In a discussion of Gödel's argument from "Is Mathematics Syntax of Language?", Lavers maintains that while in *Logical Syntax* Carnap wanted to use purely syntactic methods and abandon all use of mathematical intuition, his attitude towards the latter notion changed. According to Lavers, in the 1950s Carnap is happy to rely on an intuitive understanding of mathematics:

> [...] our intuitive understanding of the domains of number theory, or set theory, now plays the role of the explicandum in giving a systematic account of number or set. This last point is important since Gödel's arguments in these drafts presuppose that our intuitive understanding of the domain of numbers should play no role in a syntactic philosophy of mathematics. (Lavers 2019: 240)

According to Lavers, this change in position has important consequences for the matter under discussion. While he agrees that neither our intuitive conception nor formal definitions can deliver determinate truth values for *all* mathematical sentences, he thinks that in *some* cases the intuitive meaning of natural language expressions can settle their truth value in a way a formalised system cannot:

> What is defined in natural language is vague. However, this vagueness is not so pervasive that any interpretation of the vocabulary is equally acceptable. If some structure is clearly a non-standard interpretation of arithmetic, then it is clearly not this structure that we intend when we speak of the natural numbers in English. [...] If the English-language expressions concerning the natural numbers or the power-set operation are vague, then there must be some set of sentences whose truth conditions are left undetermined by the meanings of these concepts. *This class will be narrower than the class of sentences left undetermined by some formal definition of these concepts (e.g., their implicit definition by first-order axioms).* (Lavers 2008: 21f, my emphasis)

As I understand it, the view Lavers recommends is as follows: for some highly abstract claims – he mentions the continuum hypothesis – Carnap's approach indeed leads to mathematical indeterminacy, since neither the formal theory ZFC nor our intuitive conception of sets commits us either way. But this indeterminacy is not so widespread as to apply to *every* undecidable sentence. For arithmetic, for instance, there is a well-entrenched intuitive conception of the natural number sequence present in our practice of mathematics. In virtue of this, sentences like Con_{PA} can therefore have determinate truth values despite being undecidable.

Lavers refers to models that are *clearly* non-standard, and it is not so obvious what this amounts to. A model which only contains even numbers, for instance, is surely non-standard in some sense, but may not be clearly non-standard if it makes all the same sentences true as the standard model. I will here make the assumption that a model is clearly non-standard if it is not Π_1-sound, i.e. if it assigns a different truth value to some Π_1-sentence than the standard model.[7] This can be motivated in the following way: consistency sentences such as Con_{PA} are Π_1-sentences. If we want to avoid radical conventionalism and get determinate facts about syntax, we should at least exclude all models that get the facts of syntax wrong. Hence Lavers' middle way is only effective if the models excluded by our intuitive conception of numbers include the ones which are not Π_1-sound.

I think this position would be an attractive one. It is maybe not quite as ambitious as Carnap's own, since one still cannot get absolute analyticity for strong mathematical systems such as set theory. But accounting for the analyticity of arithmetic would be a major success nevertheless, given that we need determinate truth in arithmetic in order to have determinate facts about syntax. The problem with this proposal, however, is that it is unclear how exactly our intuitive conception mathematics could manage to be more powerful than a formal theory in the way Lavers describes, for reasons I will explain now.

To begin with, one might worry that it is implausible to read Carnap as relying on any kind of intuition at all. For if one interprets intution as some mysterious faculty by means of which we can acquire insights into a realm of abstract objects, then it is clear that, even in the 1950s, Carnap would not countenance anything of this kind. This is not what Lavers has in mind though. He calls our intuitive understanding of the natural number series a 'psychological fact' (Lavers 2019: 239), and there is indeed textual evidence that Carnap accepts

[7] Π_1-sentences have the form $\forall x_1 \forall x_2 ... \phi$, i.e. an arbitrary number of universal quantifiers followed by a quantifier-free formula.

intuition understood as an *ability* to make certain distinctions. In 1965 he gave a talk called "Inductive Logic and Inductive Intuition" at a conference in London, which was later published together with the ensuing discussion:

> In order to learn inductive reasoning, we must have what I call the ability of inductive intuition. [...] Maybe you have the feeling that this mysterious inductive intuition is a rather dubious basis for such a serious enterprise as the construction of a system of inductive logic. If so, I would like to call your attention to the fact that the situation in deductive logic is exactly the same. If you had before you a person who were deductively blind, that is to say, unable to distinguish between valid and invalid deductive inferences, or between deductively valid and invalid statements, even in the simplest cases, then you could not do anything with him. (Carnap 1968: 265)

Carnap illustrates the idea of deductive blindness by considering a person who is unwilling to reason in accordance with modus ponens, and can also not be persuaded to do so. Understood in this way the notion of intuition is surely a sensible one, but the use of the term 'intuition' nevertheless irritated and confused some of the philosophers in the audience, presumably since they associated this notion with the kinds of speculative philosophy Carnap rejected. Karl Popper complained about Carnap's reliance on intuition in his response to the talk, which prompted Carnap to clarify once more that there is nothing spooky going on:

> Since Popper seems allergic to the terms 'inductive' and 'intuition', let us for the moment use neutral terms. Instead of 'inductive' we shall say 'probabilistic' [...]; and instead of 'intuition': 'the ability to discriminate (at least in simple cases) between valid and invalid reasoning', or briefly 'discriminative ability'. I can hardly see reasons for serious doubt when I make the following assertion which indeed seems to me rather trivial: we cannot hope to teach someone to use a language and, furthermore, to speak and think in a logically correct way unless he has from the beginning certain basic discriminative abilities; and this holds for the whole field of reasoning, including deductive and probabilistic reasoning. (Carnap 1968: 310)

The question thus becomes whether Carnapian intuition, understood as a discriminative ability, can achieve what Lavers thinks it can: namely to pin down a conception of the natural numbers in a way that goes beyond what a formal theory such as first-order Peano arithmetic can do. As far as I can see the answer is *no*, provided that we accept the earlier assumption that human abilities need to be describable by means of a recursive theory. For since there is no way to recursively enumerate all Π_1-sentences that are true in the standard model, it cannot be that our recursively describable dispositions exclude all non-standard models which get the facts of syntax wrong. And given that the example Carnap uses to motivate his acceptance of intuition involves recursive inference rules – namely modus ponens – it is very doubtful whether he would have made room for the possibility of discriminatory capacities so powerful that they outstrip a recursive formalisation.[8]

In the end I therefore think that the cognitive constraint on human capacities – i.e. that they cannot outrun a Turing-machine – stands in the way of making Lavers' middle way a viable option. His claim that Carnap's approach to mathematics underwent a major change from the 1930s to the 1950s strikes me as insufficiently motivated as well, for while Carnap's *terminology* certainly changed, it is hard to believe that in the 1930s he would have been opposed to the idea of basic discriminative abilities to use logical inferences. But even if there was this alleged change in view, Beth's argument from non-standard models remains a stumbling block.

[8] Another way to put my criticism would be as follows: without more information on the nature of Carnapian intuition, there is good reason to be skeptical whether it can really solve the problem. The same also applies to a somewhat related proposal in the literature suggested by Sahotra Sarkar. His defence against Beth is that for Carnap, "not all models had the same epistemological status: the model with the intended interpretation was epistemologically privileged. This was the model obtained by an abstraction$_1$ from the given (typically empirical) context" (Sarkar 2013: 367). While Sarkar talks about various kinds of abstraction in some detail, it remains unclear in what sense exactly we are inhabiting an empirical context that is rich enough to abstract the standard model of arithmetic from.

Part III

Beyond *Logical Syntax*

Chapter 7

Lines of Defence

In this chapter I compare Carnap's position to some contemporary and historically influential views in the philosophy of mathematics, with a view to the question whether a response to Beth can be given by drawing on the ideas of other philosophers. We will first consider positions that, like Carnap's own, try to avoid a commitment to robust forms of Platonism, such as mathematical fictionalism. I will argue, however, that the relevant authors only avoid Beth's problem by being less ambitious than Carnap in their avoidance of external mathematical facts. Secondly, I will consider whether using logical resources that go beyond first-order logic might be of help to Carnap. The results will be negative as well, as the problem of non-recursive dispositions undermines the two most salient proposals. I then discuss a recent defence of infinitary reasoning, which if successful would indeed help Carnap against Beth. As I argue, however, Carnap's theory of dispositions prevents him from accounting for the abilities this kind of infinitary reasoning relies on. The chapters concludes with a brief look at Frege's logicism, which can be read in a way that neutralises Beth-style arguments. The costs of this seem too high for Carnap, however: the principle of tolerance needs to be given up completely, and the empiricist credentials of the Fregean position are doubtful as well.

7.1 Internalism and Pluralism

Let us begin by considering a contemporary position that can justly be described as *neo-Carnapian*: Thomas Hofweber's *internalism* about arithmetic (Hofweber 2016). Hofweber introduces a general distinction between *external* and *internal* quantification. External quantification is in effect ordinary objectual quantification, where the truth of the quantified statement '$\exists x Fx$' requires that there is some object in the domain of quantification which satisifies the predicate F. Hofweber calls the view that the quantifiers are read externally in some area of discourse externalism about this domain, and one area where externalism is plausible is talk about spatiotemporal objects.

Internal quantifiers, on the other hand, function broadly substitutionally. On the internal reading, '$\exists xFx$' is true if and only if there is a true substitution instance 'Fa'. For this to be interestingly different from external quantification it of course needs to be possible for sentences such as 'Fa' to be true even though there is no domain of quantification which contains an object that the constant 'a' refers to. And Hofweber thinks that this is indeed so: He for instance assumes that 'Paul is looking for the largest prime number' can be true, and also its existential generalisation 'There is something Paul is looking for', even though the expression 'the largest prime number' doesn't denote any object (Hofweber 2016: 150). The quantified claim thus couldn't be true on an external reading, but can be on the internal reading as no condition is put on the domain by it.

Hofweber argues in detail that internalism is true for arithmetic discourse (Hofweber 2016: chapters 5-6). And this results in a position that is recognisably Carnapian, for according to internalism the truth of arithmetic does not come with a commitment to an external domain of mathematical objects, and – so the hope – this will avoid the tricky metaphysical and epistemological questions that plague Platonists. As I will explain now, however, the problem Beth raised for Carnap will reappear as well.

We just saw that Hofweber is happy to say that using internalist quantification we can truly state that there is such a thing as the largest prime number. On the face of it this looks like a big problem. For if internalism about quantification is supposed to apply to arithmetic as well, and it is a truth of arithmetic that there *isn't* such a prime number, then we face the threat of inconsistency. Hofweber's response is an appeal to an internalist form of *quantifier domain restriction*:

> What this shows is that quantification in arithmetic can't just be *unrestricted* internal quantification. [...] [In] arithmetic we, in general, generalize over the instances that are formed with the natural number terms: 1,2,3, and so on. That the largest prime is the largest prime might be trivial, on an innocent reading of it, and thus that there is something which is the largest prime (namely the largest prime). But whether or not one of the numbers is the largest prime is so far left open. And that is the question of whether or not 1 is the largest

> prime, or 2 is the largest prime, and so on. And here the answer is no. So, on this, restricted, but internal, reading of the quantifier we get that there is no largest prime. (Hofweber 2016: 151)

That arithmetical quantification is restricted in this way is of course also crucial for the truth of claims about syntax, such as Con_{PA}. For there are non-standard numbers which can make $\neg Con_{PA}$ true, and they consequently need to be excluded from consideration as well if we want determinate truth values for undecidable statements. So we have returned to the same problem Beth put forward: how do we manage to specify the restriction we want – i.e. how do we manage to pin down the natural numbers that are the intended range of substitution instances? And this does not seem any easier to solve for Hofweber than it was for Carnap.

As it happens Hofweber does have a proposal for how to account for facts about syntax, especially consistency. He considers the view that axioms are *constitutive* of mathematical truth, i.e. to be true is to follow from some axioms. He points out that applying this model to arithmetic poses problems due to the possibility of expressing syntax within arithmetic:

> Constitutive axioms are supposed to determine all truth in their domain, but what if these axioms are inconsistent? Then no coherent domain has been established at all. That the axioms are consistent is essential for any domain with constitutive axioms. But whether the axioms are consistent is itself a mathematical question. It is, in effect, a question in arithmetic. If arithmetic, too, had constitutive axioms, then there being a well-established domain of arithmetical facts would itself depend on an arithmetical fact, leaving the domain in limbo. The fact of consistency would only obtain if it is part of the range of facts established by the axioms. But if the axioms are not consistent then no domain is established. Thus if arithmetic had constitutive axioms then none of this could get off the ground: there being a domain of arithmetical facts at all would depend on one of the arithmetical facts obtaining. (Hofweber 2016: 181)

Hofweber's response to this challenge is, in effect, to maintain that arithmetic is *special*. Whereas for other areas of mathematics it might well be that axioms are constitutive, Hofweber adopts a *logicist* account of arithmetic, according to which the Peano axioms merely partially *describe* a domain of mathematical facts that goes beyond what is settled by them:

> On the logicist picture of arithmetic defended here, axiomatizations of arithmetic, like Peano's axioms, are descriptive axioms. The axioms of arithmetic play no role in determining what is true, and they are secondary to the facts in all relevant ways. [...] The axioms don't exhaust what is true. Peano's Axioms are incomplete, like many other axiom systems, but some of the cases that are not settled by the axioms are clearly true, and not in some way indeterminate. (Hofweber 2016: 180)

One of the mathematical sentences that are clearly true despite being independent of the axioms is presumably Con_{PA}, and so the problem of accounting for consistency has solved.

This exceptionalism about arithmetic come with a major cost, however. For although Hofweber still rejects the externalist view that there is a domain of numbers that determines mathematical truth, he is now committed to objective mathematical *facts* that go beyond what can be accounted for by formal rules. And as far as I can see this move is just a rejection of internalism about arithmetic understood in *Carnap's* sense, i.e. as the claim that all truths about arithmetic follow from the rules of some linguistic framework. I will not discuss the question of why Hofweber thinks that his form of logicism is preferable to straight-out Platonism any further here. It is safe to conclude that Hofweber's aims are distinct from Carnap's though, who nowhere suggests that he wants to adopt a different treatment for arithmetic than for any other area of mathematics.

A similar lesson can be drawn from a position called *mathematical pluralism* that Justin Clarke-Doane explores in a recent book (Clarke-Doane 2020). Construed in the most radical way, mathematical pluralism is indeed similar to Carnap's internalism: it is the view that "*any (first-order) consistent* mathematical theory is true of its intended subject, independent of human minds and languages" (Clarke-Doane 2020: 160). One might think that this is just a consequence of the completeness theorem, according to which every consistent first-order theory has a model. But as Clarke-Doane clarifies, mathematical realists typically make a distinction between models, regarding some – such as the standard model of arithmetic – as intended interpretations, whereas others are unintended. Furthermore, it is commonly held that some consistent theories *only* have unintended models. We have already encountered an example of such a theory, namely PA+$\neg$Con$_{PA}$, which, from the perspective of a mathematical realist at least, only has non-standard interpretations.

The claim that every consistent theory has an intended interpretation in which it is true is thus analogous to Carnap's rejection of external question that are not pragmatic. Clarke-Doane quickly notes, however, that this radical form of mathematical pluralism leads to unwelcome results when it comes to matters of syntax:

> A model of PA + ~Con(PA) is a model in which there is an infinitely long "proof" of a contradiction from PA. I put "proof" in quotes, because a proof must be finite. The model is wrong about finiteness. Or that is that we would like to say. But if we hold that PA + Con(PA) and PA + ~Con(PA) are equally true of their intended subjects, like, say, (pure) geometry with the Parallel Postulate and geometry with its negation, then there will be no objective fact as to what counts as finite and, hence, no objective fact as to what counts as a proof in PA. Consequently, there will be no objective fact as to whether PA, or any theory which interprets it, including a regimented physical theory, is consistent! (Clarke-Doane 2020: 82f)

Clarke-Doane thinks that this kind of pluralism is too extreme, since if embraced it would, for instance, become a non-objective matter what formal theories such as PA even *are*. This is because theories are individuated by syntactic criteria, and if there is more than one correct theory of syntax there will be multiple equally correct characterisations of PA. His suggestion is therefore to moderate mathematical pluralism somewhat, such that theories which get the facts of syntax wrong do not have an intended interpretation. This is implemented by assuming that Π_1 sentences express objective mathematical facts, and then expressing moderate mathematical pluralism as the view that every Π_1-sound theory has an intended interpretation. Since consistency sentences such as Con_{PA} are Π_1 sentences, this strategy successfully excludes the unwanted deviant theories of syntax.

Clarke-Doane's treatment thus mirrors Beth's criticism of Carnap as I have interpreted it in chapter 4. The pluralism enshrined in the principle of tolerance and the rejection of factual external questions is an attractive position, but needs to be restricted when it comes to syntax. While Beth put this in terms of ontology, Clarke-Doane – similar to Hofweber – speaks of a set of objective mathematical truths (Clarke-Doane 2020: 161). But the upshot is the same. If we don't want the unattractive conclusion that there are no determinate facts about syntax, we need some resources that are theory-external:

> [...] only theories that are right about finiteness and consistency will

> count as true of their intended subjects. (Note that Con(PA), Con(ZF), and so forth are all Π_1 sentences.) This implies that the set of "objective" mathematical truths is no longer recursively enumerable, as it would be if the "pluriverse" witnessed every (first-order) consistent mathematical theory whatever. (Clarke-Doane 2020: 161)

Clarke-Doane is thus forced to postulate the existence of a non-recursively enumerable realm of mathematical truths, where these truths must be taken as basic. And in an important way this approach is not too different from the route Beth himself recommends, namely that in order to get a plausible conception of syntax we need *some* language-external resources of the kind Carnap wants to reject across the board. Beth put this requirement in ontological terms, but postulating facts or primitive truths does an equally good job. Just like Hofweber, Clarke-Doane's moderate form of pluralism thus avoids Beth's problem by restricting the internalist approach to parts of mathematics that are not arithmetic, and hence it does not constitute a defence of full-blown internalism as Carnap understood it.[1]

7.2 Fictionalism and Modality

One theme of the previous section was that while Hofweber and Clarke-Doane try to make do without a mathematical *ontology*, they cannot abstain from postulating objective mathematical *facts*, which from Carnap's perspective will seem equally problematic.[2] It is worth to dwell on the idea of *replacing ontology with ideology* some more, however, in order to evaluate a popular strategy in post-Carnapian philosophy of mathematics that relies on *modality*. Since Carnap, unlike Quine, was not opposed to using modal notions such as necessity and possibility, it makes sense to consider whether modal resources might be able to save his internalism after all.

[1] Another contemporary position that is worth mentioning here is Agustín Rayo's *trivialism* about mathematics (Rayo 2013), since Rayo himself is happy to label his view as Carnapian (Rayo 2014: 533, in response to Hofweber 2014: 443). While there has been some discussion of whether Rayo can account for the determinacy of set-theoretic claims such as the continuum hypothesis (Sider 2014, Rayo 2014: 504), in his published work Rayo does not explicitly address the determinacy of arithmetic in any detail. At this point I will therefore not investigate whether Beth's objection applies to trivialism as well, since this would involve too much guesswork.

[2] The general question of how crucial a role *objects* play in Platonistic accounts of mathematics has received some discussion (Linnebo 2018: section 1.4), see for instance *Kreisel's dictum*: "the problem is not the existence of mathematical objects but the objectivity of mathematical statements" (Dummett 1978: xxxviii).

This question is best investigated by comparing Carnap's position to a contemporary anti-Platonist position: mathematical fictionalism, as can be found in the works of Hartry Field and Mary Leng (Field 1980, Field 1988, Leng 2007). While there are various differences between their views, Field and Leng both want to account for the usefulness of mathematical discourse without invoking the existence of mathematical objects. And they also both face a particularly pressing problem: namely that in their attempts to make sense of mathematical discourse, they need to talk about formal theories and their properties. Field, for instance, thinks that although mathematical theories are strictly speaking false, they are nevertheless *useful*, and their falsity doesn't cause any damage since mathematics *conservatively extends* empirical theories. But, so an obvious objection goes, *metalogical* claims like "theory T' is a conservative extension of T" are claims about abstract objects. It is therefore tempting to conclude that the mathematical fictionalist cannot dispense with *all* mathematical entities after all, but at least needs to be a realist about syntax (Leng 2007: 90).

Once again, this line of thought is very similar to my reading of Beth's claims about the need for ontology. Field and Leng respond to it in a way we haven't examined yet, however. For while they agree that we need *some* additional resources to account for metalogical claims, they do not reach for ontology. Instead they rely on a *primitive* notion of *logical possibility* and *necessity* to explain the truth-conditions of metalogical claims. Consistency, for instance, is spelled out as follows:

> A theory T is consistent iff it is logically possible that the axioms of T are all true.

The consistency of PA is then identified with the possible truth of the axioms of PA ($\Diamond Ax_{PA}$). Note that because of the induction schema first-order PA has infinitely many axions, so in addition to the possibility operator the fictionalist will have to allow for infinite conjunctions as well (Leng 2007: 91).

Much of the discussion concerning this strategy has revolved around the question whether the move from ontology to ideology has any *epistemological* advantages, but this is not an issue I want to look into here. In our context the interesting question is whether Carnap could use the modal strategy to ensure the determinacy of syntax in a way that is consistent with his internalism.

To cut the suspense, I think the answer is 'no'. For while it is of course true that Carnap was very interested in modal logic, and even developed a sophisticated semantics for quantified model logic in his *Meaning and Necessity* (Carnap

1956b), he is not in a position to use these modal resources in the same way as Field and Leng. In a nutshell, the problem is that Carnap essentially has a reductive conception of modality, where the reduction base are once again linguistic rules, and therefore has no room for a *primitive* notion of logical possibility and necessity. In the following I will spell this out further.

In many ways Carnap's modal semantics is quite similar to the contemporary Kripkean possible worlds semantics. There is one crucial difference though. In a Kripke-semantics every frame comes with a set of worlds and an accessibility relation on the worlds, and – importantly – what this set and this relation are like is not determined by the non-modal part of the logic. This is very different for Carnap: in his system the role of worlds is played by state-descriptions, which are maximal-consistent sets of sentences, and any non-modal language automatically determines a set of state-descriptions. Williamson sums up this difference as follows:

> The semantics of the non-modal language determines a unique set of state-descriptions. [...] In this respect, his account is much more informative than contemporary possible worlds model theories. [...] Given the semantics of the non-modal language, Carnap defines what it is for a sentence of the form ◊A or □A to be true *simpliciter*, not merely true in a given model. (Williamson 2013: 48f)

For this reason one can think of Carnap's modal semantics as a *reductive theory of modality*. The aim is to let the semantics not merely determine the truth-*conditions* of modal statements, which is what a Kripke-semantics does. Rather the truth *values* of modal statements are provided by the semantics as well, since the non-modal part of the language fixes what the state-descriptions are.

Why does this conception of modality preclude Carnap from adopting the modal approach to metalogic? The problem is not that he cannot construct a linguistic framework with modal rules in which Ax_{PA}, i.e. the (infinite) conjunction of the axioms of PA, comes out as true. The issue is rather that one can also construct a framework in which the *negation* of Ax_{PA} holds. If we don't want to conclude that PA is both consistent *and* inconsistent, we need a way to discount one of these frameworks as irrelevant. But in order to do that we need to bring in some framework-external factor, and we are therefore back to the problem we started with. Beth's conclusion that determinate facts of syntax require some external facts is thus substantiated rather than avoided.

This is not to say that Carnap *could not* in principle broaden his conception of modality, and hence accept the primitive logical modality Field and Leng rely on. But this move would be a rejection of full-blown internalism analogous to the acceptance of ontology suggested by Beth. For if there are objective facts about what is logically possible or not, it will be possible to judge some linguistic frameworks by an external standard, similarly to how an external Platonist can evaluate them relative to a Model-in-the-Sky. Contrary to the initial hope, relying on modal notions thus does not give Carnap a way to save the determinacy of syntax without weakening his internalism.

7.3 Second-Order Logic and Categoricity

In this section I consider a response to Beth that relies on stronger logical resources than we have used so far, namely second-order logic. Unfortunately for Carnap, we will see that this approach also faces the problem of non-recursive dispositions. I will briefly sketch that this result generalises to other attempts of overcoming incompleteness.

In presenting the argument I so far assumed that the theories we are considering are *first-order*. About this assumption one might now complain as follows: first, in section 2.1.3 we have already seen that Carnap is happy to employ higher-order quantifiers, so the framing in terms of first-order theories is exegetically misleading. And, secondly, this assumption is of systematic importance as well. While all recursive first-order theories of arithmetic have non-standard models, *full second-order arithmetic* is *categorical*, i.e. all of its models are isomorphic to each other, and hence this second-order theory manages to pin down the standard model. In the current context second-order theories can thus no longer be ignored.

I think these points are well-taken. On one occasion Carnap in fact discusses the need to give the Peano axioms a particular interpretation:

> Peano's axiom system, by furnishing the customary formulas of arithmetic, achieves in this field all that is to be required from the point of view of formal mathematics. However, it does not yet achieve an explication of the arithmetical terms 'one', 'two', 'plus', etc. In order to do this, an interpretation must be given for the semiformal axiom system. (Carnap 1950: 17f)

In explaining how to achieve this, Carnap then refers to Frege and Russell's logicism, i.e. the program of reducing mathematics to logic – second-order logic, that is.[3] The suggestion that for Carnap second-order resources play an essential role in securing our understanding of the standard model is thus substantiated, and so let us consider whether this move in fact avoids the problem posed by Beth.

In second-orders Peano arithmetic (which I will call PA2 from now on), the first-order induction *schema* is replaced by a single second-order induction *axiom*:

$$\forall F[[F(0) \land \forall x(F(x) \rightarrow F(Sx))] \rightarrow \forall x F(x)]$$

By *Dedekind's categoricity theorem*, all full models of PA2 are isomorphic to each other.[4] Since the standard model $\mathbb{N}$ is one of these models, there is thus no sentence of PA2 that is true in one model but false in another, and hence a case like that of Carnap and Carnap* cannot be constructed. In light of this there seems to be an obvious route for Carnap to take: if he can convincingly argue that our linguistic dispositions correspond to PA2, he seems to have established that these dispositions manage to pin down the standard model of arithmetic up to isomorphism. And since it is plausible that we use second-order quantifiers in natural language anyway, this line of argument looks promising and well-motivated.[5]

Unfortunately the situation is more complicated, as we can see by unpacking the notion of a full model. Second-order Peano arithmetic only excludes non-isomorphic non-standard models if we evaluate the second-order quantifiers with what is called a *full semantics*. Second-order quantifiers range over sets of objects, but which sets exactly? On a full semantics, the answer is that the range of the second-order quantifiers is the power set of the domain of quantification – and so, in other words, we can quantify over *arbitrary* collections of objects.

On a full semantics, the domain of objects for the first-order quantifiers thus determines the range of the second-order quantifiers. This is different on a so-called *Henkin semantics*. Here one needs to specify a second-order domain explicitly, and while this might be the full power set of the first-order domain, it can also be a proper subset of the power set. And when the second-order quantifiers of PA2 are interpreted according to a Henkin-semantics, PA2 can be interpreted in non-standard ways – just as first-order PA.

3 Already in Carnap 1939: §14, Carnap uses Frege and Russell's method to give a 'normal interpretation' to a mathematical calculus.

4 See Shapiro 1991 and Button and Walsh 2018: chapter 7 for more background on the technical details and philosophical considerations of this section.

5 See for instance Boolos 1984 for reasons to think that not all quantification in natural language is first-order, and Oliver and Smiley 2016 for much more on this topic.

This observation changes the nature of the task Carnap faces. In order to argue that our linguistic dispositions pin down the standard model of arithmetic, he would have to show that we are committed to second-order logic with a full rather than a Henkin-style semantics. And it seems that this cannot be done. For it needs to be argued our dispositions commit us to interpreting the second-order quantifiers as ranging over the full power set of the natural numbers, and this task is at least as hard as the problem we are trying to solve – namely explaining our commitment to the standard numbers. This can be seen as follows: as a consequence of Dedekind's categoricity theorem, every purely arithmetical sentence is either true or false in PA2 given the full semantics. But on the assumption that our linguistic dispositions correspond to a recursive theory, it follows from the incompleteness theorems that for some arithmetical sentence we are neither committed to it or its negation. There must therefore be a gap between our recursive dispositions and what full PA2 commits one to, and hence it is implausible to maintain that we actually work within the full semantics.

For this reason it is widely held that Dedekind's categoricity theorem is not as useful for philosophical purposes as it might seem at first (Button and Walsh 2018: section 7.7). Here is a clear statement of this point by Hilary Putnam:

> Some have proposed that *second-order* formalizations are the solution, at least for mathematics; but 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". (Putnam 1980: 481)

One might think that one can get around this problem by using Dedekind's categoricity theorem in a more indirect way. It has been established that the theorem can be proved in (even quite weak versions of) first-order set theory (Simpson and Yokoyama 2013). Since such theories are recursively axomatizable there is no deep problem about what it could mean to adopt them in practice. Would this then not enable us to pin down the standard model of arithmetic after all, by relying on Dedekind's theorem from within set theory theory?

This approach will not work either. For the advantage of relying on a weak theory – namely that it is easy to argue that we can actually use it – is at the same time a disadvantage – the theory will also have non-standard models. For this reason Charles Parsons concludes that, even when used in this indirect way, Dedekind's theorem fails to deliver the philosophical cash-value we were hoping for:

> Thus, of whatever set theory in which we have proved Dedekind's theorem, there will also be nonisomorphic models. And nonisomorphic models of set theory can give rise to nonisomorphic models of arithmetic. Consider now two models $\mathbf{M}_1$ and $\mathbf{M}_2$ of set theory, and let ω_1 and ω_2 be their sets of natural numbers. [...] [Dedekind's theorem] does not tell us that ω_1 is isomorphic to ω_2; indeed, as non–well-founded models of set theory can be constructed [...] they need not be isomorphic. (Parsons 2007: 274)

Moving from first- to second-order arithmetic therefore does not provide an easy way out for Carnap. And the problem generalises to other technical means that have been studied in relation to incompleteness. Bernd Buldt's off-puttingly named paper "On RC 102-43-14" (Buldt 2004), for instance, is a very illuminating study of Carnap's attitude towards the ω-rule. Based on unpublished notes Buldt shows that Carnap was initially skeptical about Hilbert's attempt to overcome incompleteness by using the ω-rule:

> Concerning *Hilbert's new rule of inference*.
>
> *Me*: It seems to me that it does not yield more or less than the rule of complete induction; therefore, merely a question of expediency.
>
> *Gödel:* But Hilbert conceives of it differently, more broadly; the condition is meant to be the following: "If ... is provable with any metamathematical means whatsoever," and not "If ... is provable with such and such means of a formalized metamathematics."
>
> (Carnap, note from 12 July 1931, see Buldt 2004: 235)

It is not obvious why Carnap at first thought that the ω-rule is merely a different way of stating the induction schema, but Buldt conjectures that he assumed it to be a version of what is now called the *recursive* ω-rule, which instead of infinitely many premises just requires the arithmetised version of the claim that 'ϕ' is provable for all numerals:

Recursive ω-Rule

$$\frac{PA \vdash \forall n(Pr_{PA}\ulcorner \phi(n) \urcorner)}{PA \vdash \forall x \phi(x)}$$

And one might think that this proposal is interesting for more than historical reasons. The effects of iterated additions of the recursive ω-rule were further studied in the 1930s by Rosser (Rosser 1937), a paper which Carnap reviewed (Carnap 1938). And the idea of minimising incompleteness in this way has generated much more interest since the early 1960s, thanks to Feferman's work on *reflection principles* (Feferman 1991).[6] It is therefore temping to wonder whether, by using the recursive ω-rule, Carnap could dodge my objection from section 4.2.1: namely that it is impossible for human beings to follow the version of the ω-rule under discussion there, which was infinitary and non-recursive.

Unfortunately for Carnap this question has to be answered in the negative. For applying the recursive ω-rule finitely many times does not lead very far, since only a transfinite number of applications of it leads to a "considerable gain in completeness" (Buldt 2004: 241fn20).[7] The problem of how we finite human beings could actually employ the rule in an interesting way thus reappears in a different guise, and is possibly even more severe, as Carnap himself seems to assume that applying rules must be a finite procedure:

> A rule of transformation is called *transfinite* if it refers to an infinite number of premisses. (Carnap 1939: 23)
>
> We must [lay down rules] in such a way that this process of successive reference comes to an end in a finite number of steps. (Carnap 1937a: 106)

Whether allowing infinitely many premises but prohibiting infinitely many applications is a well-motivated stance can be questioned, but I will not go further into this here. For our purposes it suffices to note that there is no way around incompleteness without relying on non-recursive means, and as it stands there is no reason to think that the options considered in this section are any less problematic than the ones I already criticised.

7.4 Infinite Reasoning Regained?

Since section 4.2.1 the argument against Carnap's reliance on infinitary inference rules has been based on the assumption that using such rules in practice is

6 See Buldt 2004: 241fn20 for the equivalence between adding the recursive ω-rule and a much-used reflection principle.

7 Spelling this out more precisely would require the introduction of various technical notions, and I will refrain from doing so here so as not to loose track of the main argument. For more details on the relevant technical results and their philosophical implications see Franzén 2004, Buldt 2014, Shapiro 2019, and Wrigley 2019.

impossible. This was motivated by the naturalistic assumption that human behaviour needs to be simulable by a Turing-machine, which is taken to entail that human beings cannot follow non-recursive rules like the ω-rule. As I said there this assumption is widespread in the relevant literature, but it has recently been challenged by Jared Warren (Warren 2020b, Warren forthcoming). Against the orthodoxy he argues that human beings are able to follow at least some infinitary rules *without* postulating mysterious non-recursive abilities, which makes the prospects for certain forms of mathematical conventionalism much more rosy. Let us therefore consider whether Warren's account can be used to defend Carnap against Beth.

Warren's defence of infinitary reasoning has a negative and a positive part. To begin with, he challenges the effectiveness of two arguments that seem to underly the widespread rejection of infinitary rules. The first of these is the argument from *infinitely long proofs*: if reasoning in accord with an infinitary rule would necessarily require the construction of a formal proof, then infinitary rules can plausible be rejected, since as finite beings in a finite world we are not able to complete such a construction. But Warren argues that this requirement is too demanding, since it is not in general the case that reasoning always proceeds by way of constructing proofs (Warren forthcoming: 8). Secondly, there is the argument from *computation*: since, for instance, PA + ω-rule is an arithmetically complete theory, it may seem that if we could follow the ω-rule, we would then be able to enumerate all the truths of arithmetic. And since the latter is impossible, it must also be impossible to follow the ω-rule. Warren draws a different conclusion, however. He sees no reason to think that being able to follow the ω-rule in itself entails that we can enumerate the truths of arithmetic, since even for finite inference rules it is not the case that we can enumerate all the consequences that follow from the rules we accept (Warren forthcoming: 9). So far infinitary reasoning thus remains unscathed.

Even if it is true that extant arguments against infinitary reasoning miss their mark, however, it cannot be denied that there remains a very important difference between finitary and infinitary inference rules. While there are uncontroversial cases in which a person *actually uses* a rule like modus ponens in order to reach a conclusion – maybe by constructing a formal proof – clear-cut examples of this kind are hard to come by for infinitary rules. Since this difference in itself casts doubt on the notion of following infinitary rules, Warren gives a positive proposal of what infinitary reasoning amounts to, which we will turn to now. The guiding idea is that reasoning should be understood as a transition between mental states, which are called premise and conclusion attitudes, that fulfils certain conditions. Warren identifies three such conditions, but for our purposes it suffices to just look at the first, which involves the obtaining of certain *counterfactuals*:

> If Sally were asked for her reasons for her conclusion attitude, she would cite the content of her premise attitudes (Warren forthcoming: 11)

Furthermore, Warren claims that this counterfactual condition can in fact be fulfilled for inferences that involve infinitely many premises. At first sight this seems implausible, since presumably citing the content of an infinitary premise attitude is an impossible task. Warren therefore clarifies in what precise sense he thinks this to be possible:

> [...] Sally can have the disposition to cite as many premises as might be required. For no finite k does Sally have only the dispositions to cite the first k premises and then stop. Her dispositions always determine the next premise to be cited, after she has cited k, and this is enough. (Warren forthcoming: 14)

The following picture thus emerges: Sally can accept a statement of the form $\forall x\phi(x)$ on the basis of infinitely many premises $\phi(0)$, $\phi(1)$, ... if she has what may be described as an opened-ended disposition to cite more and more premises of the form $\phi(n)$ when asked to give reasons for accepting the universal generalisation.[8]

Warren's proposal is certainly an intriguing one, and very exciting if successful. I will not be able to further discuss and assess his defence of mathematical conventionalism, however, and instead focus on a narrower question pertinent to the aims of this thesis, namely whether *Carnap* could easily follow Warren's lead in order to deal with Beth's problem. I think that he cannot, for as I will explain now Carnap's account of dispositions seems unable to accommodate the kind of dispositions Warren's strategy requires.

[8] One might think that the problem of non-standard models reappears at this point, since *describing* the infinitely many premises $\phi(0)$, $\phi(1)$, ... presupposes a conception of what the standard numerals are. But Warren has a response to this worry: what is explanatory for him is the (alleged) fact that Sally *actually* follows the ω-rule, not the claim that she accepts some theory which includes a statement of the ω-rule (Warren 2020b: 12.III). There is thus no point at which Sally would first have to pin down the standard conception of numerals before she begins to follow the ω-rule.

In section 4.1.3 we already saw that Carnap wants to make sense of dispositions within his empiricist framework by means of so-called reduction sentences. For fragility, for instance, the following is one of the most salient reduction sentences:

(R) $\forall x \forall t$(x is dropped at $t \rightarrow$ (x is fragile $\leftrightarrow$ x breaks at t))

Earlier I already alluded to a problem this kind of view faces, namely the status of objects that are never in fact dropped. In itself (R) leaves open whether they are fragile or not, but presumably we nevertheless want to say that *some* of the undropped things are fragile whereas others are not. The most promising way to achieve this goes as follows: a glass that is never dropped nevertheless counts as fragile if it is *similar* in relevant ways to other things that are dropped and break (Quine 1969a). Spelling out the notion of similarity at play here is of course no easy task, but it does seem plausible that the extension of 'is fragile' can be made to approximate our intuitive conception in this way.

Not all problems for Carnap's account are overcome by this strategy, however, as Hugh Mellor points out:

> [N]o sense is given to fragility unless something breaks at some time. Unless some glasses are dropped and some (but not all) break, there is nothing for others to be relevantly similar to. No doubt the condition is satisfied in fact, but it is surely not necessary to a glass's being fragile. (Mellor 1974: 166)

To illustrate the importance of this point, Mellor introduces the example of an explosive fuel used in nuclear power stations. If the safety measures in such an environment are effective, the fuel will never in fact explode, and ideally this will be the case in all other power plants as well. It would be bizarre to conclude from this, however, that the fuel is not in fact explosive after all, which is what Carnap's account suggests.

This shortcoming, so I think, also makes Warren's strategy unavailable to Carnap. For as we saw above Warren needs there to be certain dispositions for which it is extremely unlikely that they will ever be manifested. Take again Sally's inference from $\phi(0)$, $\phi(1)$, ... to $\forall x \phi(x)$, and consider some arbitrarily large number a. According to Warren Sally has the disposition to justify the conclusion $\forall x \phi(x)$ by citing $\phi(a)$ if all the previous $a - 1$ premises have already been given. If a is large enough, however, citing all $a - 1$ premises may take longer than a human lifespan, and so one can be sure that the scenario never actually occurs. But if no human being ever actually cites premise $\phi(a)$ in support of the conclusion $\forall x \phi(x)$, then – just as in the case of the explosive liquid – Carnap does not appear to be in a position to claim that anyone is *disposed* to do so either.[9]

Warren himself has a response to this kind of worry. Broadly speaking his solution is to hold that the infinitely many dispositions to accept $\phi(0)$, $\phi(1)$, $\phi(2)$, ... are not independent of each other, but that there is a sense in which the disposition to accept $\phi(n)$ for arbitrary n is the result of simpler dispositions. According to him some dispositions can thus be *composed* of others:

> *S* has a composite disposition to ϕ in situation *C* iff ϕing is (or would be) the output of the iterated application of *S*'s linked simple (or complex) dispositions, in *C* (Warren 2020a: 264)

And while this move may very well result in a plausible picture of the nature of linguistic dispositions, I think that Carnap is barred from adopting the idea that dispositions can be linked in order to generate composite dispositions. For while Warren tries to remain as uncommitted on the metaphysical status of dispositions as possible, and wants his account to be neutral with regard to various competing reductionist theses (Warren 2020a: 283fn6, Warren 2020b: 34), I think that his conception of dispositions is nevertheless more robustly realist than that of Carnap. For on Carnap's approach dispositions are not a kind of entity at all, and so it is impossible to make any substantive claims about their internal structure unless these can somehow be translated back into reduction sentences. Whether this can be done for claims about linking and complex dispositions is, however, very doubtful.

Warren's account is promising for friends of conventionalist approaches to mathematics. My suspicion is that, for the reasons just described, it will have to involve at least some divergences from Carnap's original position when it comes to dispositions and the role of modality. Ultimately this may well be a price worth paying, but for now I will assume that the problem of infinitary rules persists.

---

9 The problem would not arise in this form if Carnap had a more robust understanding of modality, on which certain counterfactuals about what human beings would do can be regarded as primitive. But as I described in section 7.2 he essentially endorses a linguistic conception of modality on which modal claims need to flow from certain rules.

7.5 Frege and Logicism

The results of this chapter suggest the following master argument against the very idea of thinking that mathematics is analytic:

(1) Only non-recursive rules are powerful enough to generate all the true mathematical sentences.

(2) It is only legitimate to rely on non-recursive rules if one provides an explanation of how we finite human beings are able to use them.

(3) There is no plausible explanation of the kind (2) requires.

(C) We should therefore not identify mathematical truth with what follows from certain rules.

Is there any way to save the analyticity of mathematics in light of this? In the following I will show that it is possible to read Frege's logicism in a way that, somewhat surprisingly, rejects premise (2) of the master argument. Comparing Frege and Carnap in this way is illuminating since it highlights some of the deep differences in their thinking about the nature of logic, and also illustrates the costs of holding on to the analyticity of mathematics.

I have already argued that Carnap is not in a position to deny (2), as on his account the question of what logical system – in the form of a linguistic framework – fits our actual usage of natural language is clearly a meaningful one. For Carnap there is a sense in which natural language is 'logically amorphous' (Ricketts 2003), and so questions concerning inference rules always apply to linguistic frameworks which are coordinated with natural language, and not immediately to natural language itself. Consequently Carnap has to face the awkward question which dispositions correspond to non-recursive rules.

Being able to talk about the coordination of natural language with different logical systems is also not an incidental component of Carnap's position, for it is necessary in order to make sense of the principle of tolerance. This point is stressed by Ricketts, who draws the important comparison to Frege:

> Frege calls the laws of logic the laws of truth (Wahrsein), and says that these laws "...are boundary stones fixed in an eternal foundation that our thinking can overflow but never displace." Behind this vivid language lies a conception of the regulative role of logic in thinking. The principles of logic are to articulate the most fundamental standards for validity and consistency in thinking. Any grasp we

> have of an array of options that we can freely choose among must draw on some perhaps tacit grasp of these principles. From the perspective of this understanding of logic's regulative role, it makes no sense to represent the adoption of a logic, as Carnap does, as a choice from "an open ocean of free possibilities." (Ricketts 2004: 190f)

I think that this difference in outlook enables Frege to give a kind of response to Beth-style considerations that is not available to Carnap. To see how this is supposed to work, consider what the aims of Fregean logicism are. Broadly speaking, the idea is to show that all truths of arithmetics are really truths of logic, and hence analytic. This immediately raises the question of what can be counted as *logic* in the first place. In the *Grundgesetze* Frege uses what we would now describe as second-order logic with the infamous *Basic Law V*, which he regarded as a logical axiom (Frege 1884, Frege 1893):

> Basic Law V
> The extension of F = the extension of G **iff** $\forall x(Fx \leftrightarrow Gx)$

In this form logicism is not viable since the resulting system is inconsistent. Contemporary neo-logicists therefore usually work with second-order logic plus *Hume's Principle* (Wright 1983, Hale and Wright 2003), which can already be found in the *Grundlagen*:

> Hume's Principle
> The number of Fs = the number of Gs **iff** the Fs and Gs are equinumerous.

A lot more could be said about the technical details here, but even based on this rough sketch it would seem that the master argument against Carnap also applies to Frege. For, given Gödel's incompleteness results, the claim that mathematics reduces to logic cannot mean that all truths of mathematics can be syntactically derived using *recursive* inference rules. It therefore seems that the logicist needs to rely on something like the consequence relation of the full semantics for second-order logic (Rayo 2007). And since I argued in section 7.3 that Carnap is unable to explain how we can be committed to the full semantics, why should Frege be any better off?

To see why Frege might be in a stronger position after all, it will be helpful to introduce a distinction made by Jean von Heijenoort: that between logic conceived as a *calculus* and logic conceived as a *language* (van Heijenoort 1967).

Nowadays the former conception is probably more natural, since working on logic usually involves dealing with *metalogical* results like the incompleteness theorems. In such metalogical investigations the objects of study are certain formal theories which, for the purpose of the investigation, are regarded as uninterpreted systems of symbols. To think of logic as a language, on the other hand, is to conceive of it not as an *object* of study, but rather as the *medium* in which any kind of investigation takes place. And if, as Frege does, one accepts the idea that logic in this sense captures the most basic laws of thinking, the language of logic seems to differ from natural languages like English in at least one essential way: while one could stop speaking English and start speaking German, one cannot stop to use logic in the same way. For a non-logical language would presumably not be able to express any meaningful thoughts at all, and would hence be just gibberish rather than a genuine language. It is therefore common to say that logic understood in this sense is *universal*.

One especially salient consideration is the extent to which discourse *about logic* is possible for someone who accepts a universalist conception of logic. Van Heijenoort describes Frege's position on this question as follows:

> Another important consequence of the universality of logic is that nothing can be, or has to be, said outside of the system. [...] Frege is indeed fully aware that any formal system requires rules that are not expressed in the system; but these rules are void of any intuitive logic; they are 'rules for the use of our signs'. In such a manipulation of signs, from which any argumentative logic has been squeezed out, Frege sees precisely the advantage of a formal system. (van Heijenoort 1967: 326)

This specific formulation is slightly puzzling, since it seems that there cannot even be rules for the manipulation of meaningless symbols without presupposing *any* kind of intuitive logic. But the general idea expressed, namely that Frege has no place for metalogical enquiries, has been quite influential.

Probably the most well-known interpreter who reads Frege's conception of logic as being in tension with metalogic is Thomas Ricketts, who defended this view in a number of articles beginning in the 1980s (Ricketts 1986a, Ricketts 1985, Ricketts 1986b, Ricketts 1997).[10] According to Ricketts, Frege's rejection of a metaperspective is quite drastic, and for instance makes his position incompatible with any kind of formal semantics. Various commentators have criticised

[10] See also Goldfarb 1979 and Dreben and van Heijenoort 1986

this reading as unmotivated and too extreme (Stanley 1996, Sullivan 2004, Sullivan 2005, Heck 2007), and I do not want to commit myself to any exegetical claims here. Let us suppose, however, that Ricketts is right about one thing: namely that Frege has no place for *some* foundational enquiries in the philosophy of language, in particular those that concern the question how language makes contact with the world. If this is correct, so I will argue now, then Frege does have a way to avoid Beth-style worries.

Ricketts describes what he considers to be Frege's view on the relationship between language and reality as follows:

> [... His] view of judgment commits Frege to taking the statements of language more or less at face value. There is no standpoint from which to ask whether the thoughts expressed by the statements of language really represent reality, whether they are really true or false. Similarly, there is no standpoint from which to ask whether the statements of language really do express thoughts. (Ricketts 1985: 8, see also Ricketts 1986a: 66)

As I understand this, Ricketts interprets Frege as rejecting what is often called *metasemantics*: questions about why words have the meanings they do, names refer to what they refer to, and the like. This is supposed to be the case because, thanks to Frege's universalism, there is no standpoint from which such questions could be meaningfully raised.

I think that this idea can be generalised so as neutralise the master argument. It previously seemed that it should pose a problem for Frege as well, as he didn't seem able to answer the pressing question about our commitment to a full semantics for second-order logic either. But if Ricketts is right, then Frege would actually be able to just reject the *question itself*, since it presupposes a perspective from which our use of certain rules can be justified or questioned. Indeed, the distinction between a full and a Henkin semantics for second-order logic can only be drawn from a point of view outside of the logic we are actually using. So instead of trying to answer questions about the status of second-order logic, Ricketts' Frege is in a position to say that the relevant concerns do not even arise.

If this interpretation of Frege is coherent, there is at least some conceptual space for a view according to which mathematics is analytic and to which the objection from non-recursive dispositions does not apply. Is this good news for Carnap's project though? I think not. Setting aside the fact that Ricketts' interpretation is highly contested, Frege's general outlook is in some respects diametrically opposed to that of Carnap. Frege clearly believes that there is *one true logic*, for he calls the laws of logic "boundary stones which are anchored in an eternal ground, which our thinking may wash over but yet cannot displace" (Frege 2013: xvi). Furthermore, it is far from clear whether Frege's anti-psychologism about logic is even compatible with empiricism, and his notorious remarks on the *third realm* of thoughts are not too reassuring (Frege 1918).[11]

The Fregean defence of the analyticity of mathematics I sketched thus seems to abandon two of the most central tenets of Carnap's philosophy, namely the principle of tolerance and empiricism. Those who sympathise with Carnap are therefore probably better served by looking forwards in time at the position of Quine. While he famously gives up the analyticity of mathematics, Quine remains a committed empiricist. And even though he doesn't have anything as strong as Carnap's principle of tolerance, unlike Frege Quine at least makes room for the revisability of logic.[12] In the next chapter we will therefore investigate whether Quine's philosophy of mathematics can withstand Beth's objection.

11 In an unpublished paper, Frege for instance describes grasping a though as a process in which "something comes into view whose nature is no longer mental in the proper sense [...] and this process is perhaps the most mysterious of all" (Frege 1979: 145).

12 The coherency of which is contested, however, see for instance Arnold and Shapiro 2007.

Chapter 8

Quine on Mathematical Truth

In this chapter I will outline what I take to be Quine's position on the truth of undecidable mathematical sentences. This is not an easy task, as Quine's attitude towards mathematical truth is much more complicated and subtle than an initial reading of his work may suggest. Nevertheless I will provide some reasons to think that Beth's argument does not have any force against Quine – despite the many continuities of his position with Carnap's. This is due to some fundamental differences in their respective conceptions of the nature of language and logic, with Quine being more similar to Frege in that he *rejects* certain questions Carnap needs to give answers to. My conclusions will be tentative, but will illuminate some of the deep differences between Carnap and Quine that are easily missed.

8.1 Quine and Mathematical Platonism

In the short chapter 3 we already saw that Quine rejects the idea that mathematical statements are analytic, true in virtue of meaning or rules, or anything in the vicinity. For him there is just truth *simpliciter*, and empirical and mathematical statements are true in the same (and only) way. But what is Quine's *positive* story about truth in mathematics, especially with respect to the undecidable sentences that have occupied us so much? The usual textbook and encyclopaedia story about Quine and the philosophy of mathematics goes as follows: Quine rejected Carnap's linguistic doctrine of mathematical truth, and thereby Carnap's idea that using mathematics does not come with any ontological commitments. The abstract objects mathematics would have to be about seemed mysterious, however, and for this reason Quine sympathised with nominalism. This experiment culminated in the joint paper with Nelson Goodman (Goodman and Quine

1947), but afterwards Quine came to realise that abstract entities – in particular sets – were indispensable for science, and hence – grudgingly – became a mathematical Platonist, although an unenthusiastic one who would have preferred to do without abstract entities any day.[1]

While there is certainly something correct about this story, the sense in which Quine's mature position is a form of Platonism is not at all straightforward, and can be easily misunderstood. It is tempting to think that Quine must have embraced a kind of *external* Platonism according to which what Tait described as a Model-in-the-Sky determines the truth values of mathematical sentences. For unlike Carnap, Quine seems happy to regard ontological questions as factual rather than pragmatic, as can be seen in a famous line from "Two Dogmas of Empiricism":

> Ontological questions, under this view, are on a par with questions of natural science. (Quine 1951: 43)

Since Quine definitely thinks that Carnap's attempt to account for mathematical truth in terms of linguistic rules has failed, it would thus only be natural to read him as accepting an ontology of abstract entities in order to fill the explanatory gap left by abandoning the analyticity of mathematics. On this view, undecidable mathematical sentences are true or false in virtue of there being a realm of abstract objects.

If this is correct, then Quine's attitude towards ontology would indeed be radically different from Carnap's: ontological investigations can be said to give us insights into the structure of reality, showing in this case that there must be abstract objects in addition to the concrete ones. The problem, however, is that this picture does not fit how Quine describes his own conception of the role of ontology at all:

> The scientific system, ontology and all, is a conceptual bridge *of our own making*, linking sensory stimulation to sensory stimulation. (Quine 1981a: 20, my emphasis)

[1] Somewhat surprisingly there is not that much secondary literature specifically on Quine's philosophy of mathematics. Two salient older papers are Parsons 1986, to which Quine responded (Quine 1986a), and the little-known Hart 1979. Isaacson 2004 contains an illuminating discussion which flags various tensions in Quine's position. Isaacson also mentions that he plans to discuss Quine's account of mathematics further in another paper (263fn25), but unfortunately this was never realised (personal communication). Burt Dreben was commissioned to write a chapter on truth for the *Cambridge Companion on Quine*, which would presumably have addressed the issue of truth in mathematics, but unluckily for us Dreben died before writing it (Dreben 2006: 287).

> [...] our ontology, like grammar, is part of *our own conceptual contribution* to our theory of the world. (Quine 1990: 36, my emphasis)

Quine's suggestion is that *we* are the ones who impose an ontology onto the world. In order to better understand what he could mean by this, it is helpful to consider his notorious doctrine of *ontological relativity*: the thesis that, roughly, the ontology of any theory is relative to a background theory (Quine 1969b). Suppose we accept a theory about the empirical world in which we have predicates for animals such as cats and dogs. If we think about such a theory in a model-theoretic way, there will be a domain of quantification the quantifiers of our theory range over, and names and predicates apply to objects in this domain as well. It is now natural to assume that the domain of an empirical theory actually contains empirical objects, such as cats and dogs. But, Quine observes, from the perspective of the object language of the theory itself it does not really matter *which* entities are in the domain at all. We can interpret the theory in a domain consisting only of numbers as well, while preserving all the truth values of the object language sentences. The sentence "there are three dogs on the lawn" can come out true, in other words, regardless of whether, from the perspective of a metalanguage, 'dogs' is interpreted as denoting dogs or numbers.

As a technical result this is almost trivial, but Quine takes it to have substantive implications. According to him, the possibility of truth-preserving permutations of the domain shows that there is no fact of the matter as to what the correct ontology of *our* best theory of the world is, since empirical evidence could neither favour or refute any claims about what the domain of quantification looks like (Quine 1981a: 22).

This brings us to the important idea that, for Quine, sentences and their truth are in a sense primary. Quine's general conception of how language relates to reality is as follows: language makes contact with the world at the level of *sentences* (or even at the level of theories, which are sets of sentences). We are interested in theories that are true, and for Quine *truth*, unlike ontology, is *not* relative (setting aside the trivial relativity of truth to a language). And we are interested in *objects* only to the extent that they are needed to account for the transition between sentences:

> It is occasion sentences that report the observations on which science rests. The scientific output is likewise sentential: true sentences, we hope, truths about nature. The objects, or values of variables, serve merely as indices along the way, and we may permute or supplant

> them as we please as long as the sentence-to-sentence structure is preserved. (Quine 1981a: 20)

From Quine's perspective it is therefore indeed conceivable, even though actually false, that we have a theory that tells us all there is to say about reality without relying on the notion of an object at all. Suppose, for instance, that our best overall theory of the world would not require quantifiers, but could be expressed in propositional logic. In such a scenario all we are dealing with are sentences and implication relations between them, and in this sense describing the world does not require postulating objects. Quine repeatedly makes this point with respect to other weak theories:

> [... In the case of a finite universe of named objects] there is no occasion for quantification, except as an inessential abbreviation; for we can expand quantifications into finite conjunctions and alternations. Variables thus disappear, and with them the question of a universe of values of variables. (Quine 1969b: 62)

One can sum this up as follows: Quine is a realist about truth but an instrumentalist about ontology. Our best theories are such that they demand a universe of objects to be true, but this only means that there need to be *some* objects or other. Simple model-theoretic considerations show that if there is one set of objects that makes a theory true, then there are many, and the upshot of ontological relativity is that no one ontology is privileged over another as long as truth values are left intact. Quine's *primacy of sentences and their truth* is thus in stark contrast with philosophers who want to explain the language-world connection by invoking something like *direct acquaintance* with particular objects, like Russell in some periods of his life.[2]

What are we to make of this? The literature on ontological relativity is immense, and opinions differ widely on how exactly this doctrine should be understood, what Quine's arguments for the thesis are, whether those are any good – and, indeed, whether the position is even coherent.[3] We will not be able to go

[2] Hylton helpfully contrasts Quine with Russell in Hylton 2006.

[3] Eklund 2007 is a useful starting point, as it contains references to many of the most important papers. Hilary Putnam, whose *internal realism* often looks quite similar to Quine's position, is actually one of the philosophers most strongly opposed to ontological relativity:

> Doubtless Quine and Dreben will reply that one's statements about objects can be 'robustly' true or false even if one's reference to those objects 'float freely'. I would discuss this claim if I could make sense of it; but, alas, I cannot. (Putnam 1992: 396)

into more details here, but we have seen enough to come back to the question from the beginning: namely whether Quine endorses a version of mathematical Platonism according to which mathematical truth is explained by relying on mathematical ontology.

I think there is reason to be skeptical about giving ontology a serious explanatory role in Quine's philosophy. For the primacy of sentences suggests that the explanatory direction should go precisely *the other way*, such that the ontology of mathematics would be determined by which mathematical sentences are true. When motivating ontological relativity, Quine after all supposes that sentences have fixed truth values irrespective of any particular interpretation, and so in this sense the truth of sentences is what determines the admissible ontologies. I would therefore hesitate to describe Quine as an external Platonist who assumes the existence in a Model-in-the-Sky. The latter move seems dangerously close to postulating a realm of *Dinge an sich* to account for empirical phenomena, and clashes with Quine's description of the way we interact with the world:

> Man proposes; the world disposes, but only by holophrastic yes-or-no verdicts on the observation sentences that embody man's predictions. (Quine 1990: 36)

Let us therefore consider whether taking the idea that truth determines ontology seriously results in a more plausible reading of Quine.[4]

8.2 Beth's Revenge?

Applied to mathematics, Quine's stress on the primacy of sentences and their truth suggests that our mathematical ontology is determined by which mathematical sentences are true, rather than the other way round. In this section I will show that this interpretation of Quine also has its problems, as it faces a version of Beth's objection. Furthermore there are passages in which Quine explicitly denies some core assumptions underlying this reading, which makes his considered position all the harder to pin down.

[4] As yet I am undecided whether there is merely a *tension* between external Platonism and ontological relativity, or a genuine *conflict*. In order to assess this more fully one would have to further investigate the relationship between Quine's position and *structuralism*, for it seems that even if a particular set of objects itself cannot play an interesting explanatory role, the *structure* all admissible ontologies of a theory share might still be able to do so. In the late paper "Structure and Nature" Quine indeed expresses sympathies for structuralism, but also adds that "my global structuralism should not [...] be seen as a structuralist ontology" (Quine 1992: 9), which can be read as cautioning against the previous proposal.

If the ontology of mathematics is determined by *any* theory, it will presumably be determined by *our* best overall theory of the world. But what theory is that exactly? This is a big question, but for our purposes it suffices to focus on the idea that the theory is *ours*. For there must be a sense in which we actually accept or are committed to the theory that determines ontology, just as for Carnap we could wonder about what *using* a framework actually means. Quine helpfully describes what it is for a theory to be someone's theory as follows:

> In *Word and Object* and related writings my use of the term 'theory' is not technical. For these purposes a man's theory on a given subject may be conceived, nearly enough, as the class of all those sentences, within some limited vocabulary appropriate to the desired subject matter, that he believes to be true. Next we may picture a theory, more generally, as an imaginary man's theory, even if held by nobody. (Quine 1969c: 309)

This should make some alarm bells ring, however. Quine enthusiastically shared Carnap's linguistic behaviourism, according to which believing a sentence to be true needs to be spelled out in terms of dispositions to act in certain ways. Therefore the argument from section 4.2.1 applies: our linguistic dispositions need to correspond to some recursive theory, and hence the set of sentences we hold to be true must be recursive as well. And this seems to give us an easy way to turn Beth's objection against Quine:

(1) Whatever our best theory of the world may look like, it must be recursive.

(2) There will hence be mathematical sentences whose truth values are not settled by the best theory.

(3) These undecidable mathematical sentences therefore have no determinate truth value.

(4) Consequently our best theory cannot pin down the standard model (or a model isomorphic to it) as its ontology.

(5) So our mathematical ontology is not isomorphic to the standard model.

If this were right, Quine would therefore face the same issues as Carnap in accounting for the truth of undecidable sentences.

As it stands, however, this argument is not successful. For the move from (2) to (3) relies on the following assumption:

(A_1) For a sentence to be true is for it to follow from the best theory of the world.

And the step from (4) to (5) relies on a second assumption:

(A_2) What there is, i.e. our ontology, is determined by the ontological commitments of the best theory of the world.

But Quine actually rejects both of these. In a paper called "Quine's Truth" (Bergström 1994), Lars Bergström suggested that Quine might be committed to a version of (A_1). Quine's response is an unambiguous denial:

> He asks, rhetorically I suspect, whether I hold that 'a sentence is true if and only if it follows logically from our theory or an idealized version of it'. No. Such an ideal theory is impossible by Gödel's theorem. [...] Or do I hold that a sentence is true 'if and only if implied by the best humanly devisable theory'. No; Gödel speaks again. (Quine 1994: 496)

Quine thus clearly thinks that there are sentences which have determinate truth values even though they are not entailed by any (recursive) theory.

As Quine's tone suggests, (A_1) presumably looked quite unappealing anyway. (A_2) may seem somewhat more plausible, as it is tempting to understand Quine's well-known notion of *ontological commitment* in a way that supports it. Here is how Quine characterises this notion in "On What There Is":

> [A] theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true. (Quine 1953: 13f)

Given this it is very natural to think that what there is must somehow be determined by the ontological commitments of some theory. But Quine explicitly cautions against this only a few pages later:

> We look to bound variables in connexion with ontology not in order to know what there is, but in order to know what a given remark or doctrine, ours or someone else's, *says* there is; and this much is quite properly a problem involving language. But what there is is another question. (Quine 1953: 15f)

The last sentence is not just a throwaway remark. In a paper in which Terence Parsons distinguishes various non-equivalent ways of spelling out the notion of ontological commitment, he stresses this point some more, obviously with Quine's approval:

> This example further shows the distinctness of Quine's notion of "ontology" from any of his notions of ontological commitment. For we have determined the ontological commitments of T_1 in all three senses without even knowing what its ontology is (without knowing the actual range of the variables). The ontological commitment of a theory fixes neither the actual nor the possible ontologies of a theory.) [Footnote: Quine has stressed the importance of this fact in understanding ontological commitment.] (Parsons 1970: 68f, 73fn12)

So Quine would clearly not have accepted (A_2) either.

As presented above, the Beth-style revenge argument therefore does not apply. But the discussion is bound to leave us in a puzzled state as to the nature of Quine's considered position on the relationship between mathematical truth and mathematical ontology. The results of this section push us towards reading Quine as accepting something like a Model-in-the-Sky after all, since he rejects attempts to explain ontology in terms of truth. But it is still not easy to see how this is compatible with the quotes we saw in the previous section, and the claim that sentences and their truth are primary. One might thus wonder whether Quine has a coherent position on the nature of mathematical truth at all.

In the next section I will try to make a positive proposal, by indicating a way to read Quine as presenting a potentially attractive alternative to Carnap. In order to do this I will set questions about ontology aside, and focus exclusively on truth. After all, the crucial question – whether Quine is in a position to hold that undecidable mathematical sentences have determinate truth values – can be asked without any mention of ontology. I think that this will lead to more illuminating results, as the confusion we encountered indicates that the whole idea of an explanatory link between truth and ontology has no place in Quine's philosophy. If this is correct, then Quine's rejection of Carnap's conventionalism can be described in the following way: it is not that we should explain mathematical truths in terms of ontology *instead of* linguistic rules, but one should rather give up the fundamental assumption that mathematical truth *requires* a special kind of explanation in the first place. Let us therefore explore whether this idea gives Quine a way to deal with undecidable sentences.

8.3 Quinean Internalism

It will be helpful to start with a big-picture question: why did Carnap want mathematics to come out as analytic in the first place? The motivation he gave in "Empiricism, Semantics, and Ontology" was that in this way the apparent tension between a commitment to empiricism and talking about abstract objects is resolved. That there is such a prima facie conflict between empiricism and the acceptance of abstract entities is an assumption Quine shares, and he likewise wants to hold on to empiricism. Instead of analyticity, however, Quine's favoured remedy is *holism*, and he helpfully describes this departure from Carnap as follows:

> How, Carnap asked, can mathematics be meaningful despite lacking empirical context? His answer was that mathematics is analytic. Holism's answer is that mathematics, insofar as applied in science, imbibes the shared empirical content of the critical masses to which it contributes. [...] Once we appreciate holism, even moderate holism, the notion of analyticity ceases to be vital to epistemology. (Quine 2008: 26f)

While much could be said about the nature of Quine's holism, it is at least reasonably clear how it can account for the meaningfulness of *some* parts of mathematics. The thought is that our best physical theories cannot be formulated without a good amount of mathematics, and so plausibly the best overall theory of the empirical world will contain something like Peano arithmetic or a version of set theory. Granting this, however, we get at most the result that mathematical sentences that follow from some recursive theory come out as meaningful. What about undecidable sentences, which are formulated in the mathematical language but independent of the relevant axioms?

In some cases it is possible to tell a story about how the truth of undecidable sentences relates to observable facts, thus securing their empiricist credentials. Take Con_{PA} again: if it were false, we would expect that at some point someone would actually derive a contradiction from PA, and hence show it to be inconsistent. Since PA is a theory that has been used and studied for decades without any contradictions turning up, it is generally believed to be consistent, and this plausibly provides some inductive evidence for Con_{PA}.[5] It must be said,

[5] Exceptions prove the rule: in 2011 the mathematician Edward Nelson claimed to have a proof of the inconsistency of PA, but this was later found to be flawed (`https://golem.ph.utexas.edu/category/2011/09/the_inconsistency_of_arithmeti.html`).

however, that Con_{PA} is an unusual case. Most of the infinitely many undecidable sentences will never be of any practical use in physical theory, and it is unclear whether we can even imagine them making any difference to the empirical world. The crucial question thus becomes whether Quine is in a position to maintain that *all* mathematical sentences are meaningful and have determinate truth values, or just the subset of them that is actually needed to account for empirical phenomena.[6]

It is not always obvious how much of mathematics Quine actually wants to come our as meaningful. In his writings he usually focusses on set theory, and describes his attitude as follows:

> So much of mathematics as is wanted for use in empirical science is for me on a par with the rest of science. Transfinite ramifications are on the same footing insofar as they come of a simplificatory rounding out, but anything further is on a par with uninterpreted systems. (Quine 1984: 788)

The picture that emerges is one where mathematics can be divided into three distinct regions. Some mathematical sentences will straightforwardly count as meaningful in virtue of being needed to do empirical science. Other mathematical sentences are not strictly speaking necessary for science, but it is nevertheless beneficial to treat them as meaningful since mathematical theories that include both sentences of the first *and* second kind are simpler than ones which only contain those of the first kind. And thirdly there are mathematical sentences which do not even play an indirect role in any empirical context, and Quine wants to treat them as uninterpreted symbols.

In which category do the undecidable sentences of arithmetic fall? I think that Quine would regard them as meaningful even though most of them can be expected to play no direct role in empirical science. This is because they are formulated in the language of arithmetic, which Quine generally regards as meaningful, and discounting them would therefore require giving up standard principles of bivalence and compositionality. He admits that countenancing undecidable sentences is a theoretical cost, however:

> We stalwarts of two-valued logic buy its sweet simplicity at no small price in respect of the harboring of undecidables. [...] The matter is undecidable, but we maintain that there is a fact of the matter. [...] on

[6] For a more in-depth discussion of Quine's naturalism about mathematics and the challenges it faces see Maddy 1997.

> the mathematical side, for the continuum hypothesis or the question of the existence of inaccessible cardinals. (Quine 1981b: 91)

Since Quine held onto classical logic until the end he clearly thought that, on balance, this is a price worth paying. I therefore think that he would justify the claim that undecidable sentences have determinate truth values in the following way: first, we need some mathematical theory to do physics in any case, and using its vocabulary we can formulate purely mathematical sentences which are undecidable. Since we are also committed to classical logic we need to accept that each of the undecidable sentences is either true or false. In many cases we will have no idea *which* truth value it is, as there is no clear connection between the mathematical sentences and empirical claims. While that is unfortunate the resulting picture is nevertheless more simple and elegant than one where bivalence is given up, since accepting that undecidable sentences have determinate truth values does no real harm either, and so reforms are not urgent.

This argument based on bivalence and simplicity will probably appear suspicious to many. It seems that Quine in effect commits himself to something like *brute facts* about mathematical matters, similar to some of the contemporary positions surveyed in section 7.1. And one might therefore think that empiricism has gone over board after all, for we end up with a realm of facts about abstract objects that are in no clear way related to anything in experience.

This is indeed a pressing worry, and Quine himself seems to have felt some uneasiness about the status of mathematical facts. He reports being pressed about this issue by Burt Dreben, but the paper from which this passage is taken contains no clear resolution on Quine's part:

> Dreben once put me a [...] challenging question: is there no fact of mathematical matters? For me, unlike Carnap, mathematics is integral to our system of the world. Its empirical support is real but remote, mediated by the empirically supported natural sciences that the mathematics serves to implement. On this score I ought to grant mathematics a fact of the matter. But how, asks Dreben, does this involve the distribution of microphysical states? What would there being a largest prime number have to do with the distribution of microphysical states? (Quine 1986b: 430)[7]

There might be a way out, however. In an interview from 1994, Martin Davies explicitly asked Quine about the status of truths in higher set theory, where even

[7] For a discussion of Quine's notion of facts of the matter that also suggests an interesting response to the current challenge see Ricketts 2011.

indirect connections to experience are impossible to make out. The following is an illuminating passage:

> *Davies*: I wanted to ask whether in the outer reaches of set theory, where you said along the way in your answer that you would advert to considerations of simplicity for example, whether its best to think of simplicity as a guide to the truth or as constituting the truth in those cases.
>
> *Quine*: Ah yes, I don't think I can distinguish that once we get that far out. Just as when we're way in at the other extreme at the most obvious, the truth functions and the like, I can't distinguish between what's a change of logical theory and what's a change of terminology and semantic interpretation. (Fara 1994: "The Dreben Panel", my transcription)

This suggests a response to the worry about Quine's commitment to brute mathematical facts. The objection relies on the assumption that mathematical truth necessarily needs to *track* some feature of the world – hence no truths without the corresponding facts. But while this is a reasonable assumption for some areas of discourse, the tracking metaphor has no sensible application in the most abstract parts of mathematics. Accepting that all mathematical sentences have determinate truth values is therefore possible without the kind of metaphysical baggage that would conflict with empiricism.

Whether this argument from bivalence and simplicity is ultimately successful requires a more extensive study which I will not be able to pursue here. Much remains unclear, for instance what it could mean to say that simplicity *constitutes* truth in some cases. But I do think that Quine's way presents a genuine alternative to Carnap's approach, and should be explored further by those attracted to an empiricist outlook.

Assuming that Quine's account has a chance of success, one obvious question is whether Carnap could not just replicate it, and hence avoid Beth's objection after all. I do not think that this can be done in a way that retains the core commitments of Carnap's position, however. For there is a sense in which Quine can be regarded as a more thoroughgoing internalist than Carnap: Quine does not draw a sharp distinction between natural languages and constructed frameworks, and is therefore happy to accept the notion of truth *simpliciter* rather than just truth in some framework. Hence Dreben's enigmatic remark that the "core of Quine's difference with Carnap over the string 'analytic"' is "Quine's insistence to Carnap that truth is truth and existence is existence" (Dreben 1996: 57fn18).[8]

Quine's considerations therefore do not really address the same question that made trouble for Carnap: namely in virtue of what we speak a linguistic framework with one definition of 'analytic' for mathematics rather than another. Indeed, for Quine this is a question that does not even arise, as for him we always start theorising from the perspective of one language or other, and can then wonder about truth and falsity within that language (Quine 1960b: 24). Carnap, on the other hand, can only start to ask such questions once it is settled what framework we are using (Carnap 1942: 14). He therefore faces a much harder task: he does not only need to invoke considerations of simplicity in order to motivate the acceptance of bivalence *in general*, but would have to argue that there is one *specific* framework for mathematics that is simpler than any other. And I do not see how this could be achieved. To be sure, Carnap might be able to emulate Quine if he were ready to treat mathematics as synthetic, and thus abandon the notion of analyticity altogether. But this would arguably be more of a capitulation than a defence of Carnap.

8 We already encountered this very important point in section 3.1. Ricketts aptly describes Carnap's conception of language as Hilbertian, whereas Quine's is more akin to Frege (Ricketts 2009.

Chapter 9

The Empirical World

In this last chapter I will venture beyond the philosophy of mathematics in order to look at Carnap's attitude towards the ontology of the *empirical world*: observable objects like tigers and theoretical entities posited by the sciences, such as electrons. My ambition is not to give a comprehensive account of Carnap's treatment of observation and empirical science. The main motivation for looking at this area at all is that Carnap takes abstract mathematical objects to play an important role even in the empirical parts of language. Another reason for ending the thesis on this note is that one general lesson that can be drawn from Beth's argument – namely that problems start to appear once one scrutinises the relationship between natural language and constructed frameworks – also applies to the empirical case. As we will see in the last section, moreover, some difficult questions Carnap faces also arise for a form of neo-Carnapianism, namely Amie Thomasson's *easy ontology*.

9.1 Hyper-Pythagoreanism

So far we have been concentrating almost exclusively on the ontology of mathematics, and hence on abstract entities such as numbers or sets. I have described Carnap's position as an *internalist* version of Platonism, because he accepts that we can truly say that *there are* such abstract entities, which goes beyond what a nominalist is willing to admit. Since the truth of mathematical existence statements is supposed to follow from the linguistic rules of a framework, however, it does not commit Carnap to metaphysical posits such as the external Platonist's Model-in-the-Sky. The stability of this position has come under pressure by Beth's argument, but let us set this worry aside in order to ask a different question: what, according to Carnap, is the status of the ontology of the empirical world, and existence statements about tigers and particles?

When I introduced Carnap's notion of a linguistic framework in section 1.1.2, we saw that the distinction between internal and external questions and statements applies to all areas of discourse, both mathematical and empirical. This means in particular that Carnap would also reject *factual external* questions about empirical objects, such as whether there *really are* any tigers, independently of a linguistic framework which provides rules that allow one to assess this question. Consequently, just as Carnap denies that there is a Model-in-the-Sky understood as a language-independent arbiter of mathematical truth, he would also reject the view that the world *in itself* provides a privileged domain of empirical objects to quantify over. Instead, each linguistic framework will have its own internal domain of quantification.

One important difference between abstract and empirical objects, however, is that in the empirical case the rules of the frameworks should not determine *which* entities exactly the domain of quantification contains. For whether there are any tigers at all, and if so then how many, is clearly not a matter of linguistic conventions, but depends on the state of the world. So while for mathematical discourse the framework rules in themselves determine what we quantify over, in the empirical case it is framework rules *plus* the world which – in some as yet unspecified way – settle the extent of the domain of quantification.

So far Carnap's views on the ontology of the empirical world seem pretty straightforward. But the situation is complicated by the fact that on a number of occasions, the picture Carnap paints is actually quite different from what I have described. In *Logical Syntax* he introduces the concept of a *coordinate language*, which, as it is first presented, appears to be a language in which objects can be referred to in an especially systematic way, for instance by means of space-time coordinates:

> A language which is concerned with the objects of any domain may designate these objects either by *proper names* or by systematic positional *co-ordinates*, that is by symbols which show the place of the objects in the system, and, thereby, their positions in relation to one another. [...] We shall call a language (or sub-language) which denotes the objects belonging to the domain with which it is concerned by positional designations, a *co-ordinate language*, in contradistinction to the *name-languages*. (Carnap 1937a: 12)

But later in the book it becomes clear that for Carnap the distinction between name and coordinate languages does not just concern *how* objects are designated, but also relates to *what* objects the quantifiers of the language range over in the first place. For in a discussion of whether the axiom of infinity can be considered a logical law despite its ontological commitments, Carnap has the following to say:

> In our object languages I and II, the matter is quite different owing to the fact that they are not *name-languages* but *coordinate-languages*. The expressions of type 0 designate not objects but positions. The axiom of infinity (see §33, 5a) and sentences like '(∃x)(x = x)' are demonstrable in Language II, as are similar sentences in language I. But the doubts previously mentioned are not relevant here. For here, those sentences only mean, respectively, that for every position there is an immediately succeeding one, and that at least one position exists. But whether or not there are objects to be found at those positions is not stated. (Carnap 1937a: 141)

Carnap seems to say here that in a coordinate language the sentence '(∃x)(x = x)' does not express the claim that an object exists, which is a bit misleading. For on Carnap's considered view, the positions that are being quantified over are *sets* – quadruples of numbers representing space-time coordinates, more specifically. In the quote therefore Carnap doesn't seem to count abstract entities as genuine objects, and one should read 'object' as shorthand for 'empirical object' in this context.[1]

Understood in this way, Carnap's preference for coordinate languages leads him to a *purely mathematical* ontology of sets and numbers, with no concrete objects such as tables and tigers showing up in the domain of quantification at all. This may seem bizarre at first, but a few decades later Quine also considered the advantages of such a position, for which he introduced the term *hyper-Pythagoreanism*. In his "Whither Physical Objects?", he – rightly, I think – traces it back the idea to Carnap's coordinate languages:

> Carnap was propounding such a *Koordinatensprache* already in 1934, and not because of constraints on the notion of physical object from the side of physics; for the scheme has also a certain intrinsic appeal. Numbers and other mathematical objects are wanted in physics anyway, so one may as well enjoy their convenience as coordinates for physical objects; and then, having come thus far, one can economize a little by dispensing with the physical objects. (Quine 1976: 502)

[1] For more on Carnap's views on the axiom of infinity see Lavers 2016b.

The natural question to ask is how hyper-Pythagoreanism can be taken seriously in light of the obvious fact that reality is not purely mathematical. I and the computer I am typing on right now are clearly neither sets nor numbers, for instance, so in what sense could it be that all objects are abstract? Quine explains why hyper-Pythagoreanism is not a completely insane view as follows:

> We must note further that this triumph of hyper-Pythagoreanism has to do with the values of the variables of quantification, and not with what we say about them. It has to do with ontology and not with ideology. The things that a theory deems there to be are the values of the theory's variables, and it is these that have been resolving themselves into numbers and kindred objects – ultimately into pure sets. The ontology of our system of the world reduces thus to the ontology of set theory, but our system of the world does not reduce to set theory; for our lexicon of predicates and functors still stands stubbornly apart. (Quine 1976: 503)

The guiding idea is that while all objects are purely mathematical, some of the *predicates* of our theory of the world are nevertheless empirical, which means that whether or not they apply cannot be decided based on purely mathematical considerations. And this approach can also be found in Carnap's writings, who explains how one can make empirical statements in a coordinate language as follows:

> [...] ‘Blue($0''$)‘ may be read: “The position with coordinate 2 is blue”. Strictly, ‘$0''$‘ designates only the pure number Two; reference to the position does not belong to the significance of ‘$0''$‘, but to that of the predicate ‘*Blue*’, whose significance is “The position having ... as coordinate is blue.” It is convenient, however, to speak as if the individual expressions designate not only numbers, but coordinated positions of the system as well. For this reason we often call such positions (be they space points, time points, or space-time points) the individuals of the coordinate language in question. (Carnap 1958: 162, §40)

Analogously, talking about tigers in a coordinate language would presumably work as follows: the statement 'there is a tiger' corresponds to the claim that there is a collection of coordinates (which are abstract but represent space-time points), such that an empirical predicate that corresponds to 'x is a tiger' applies to this collection. In a loose sense even hyper-Pythagoreanists can thus admit that there are empirical objects, they just need to interpret this claim in a way that does not commit them to tigers being in the domain of quantification.

We have now seen two ways in which Carnap can account for talk about empirical objects: what one may call the *commonsense* view on which frameworks have both abstract and concrete objects in their domain of quantification, and the hyper-Pythagorean position in which contact with the empirical world is made purely though predicates. What is the relationship between these two different options, however? Is only one objectively correct, and must we hence choose sides? Knowing Carnap as we do, it seems pretty obvious that he would say that the choice between these two ways of doing things is a pragmatic question of convenience. There are some passages, however, that point in a different direction, and suggest that it is philosophically important to think of the ontology of *observable* entities in a particular way. We will turn to this issue in the next section.

9.2 Observable and Theoretical Entities

The fact that Carnap at times proposes a hyper-Pythagorean position vividly demonstrates something I already mentioned in section 1.1.3: namely that for him, language doesn't make contact with the world by an association of names with objects. For both Carnap and Quine the distinction between objects and properties emerges from the theory-internal distinction between names and predicates, and so there is no place for a relation of direct acquaintance with entities. This is at least what is suggested by the remarks we have discussed so far, and I also think that it is ultimately the correct reading of Carnap. But there are admittedly passages which suggest that there is a part of language that is special, and obeys different rules: namely the *observation language*, in which we talk about macroscopic everyday objects. Let us therefore consider whether Carnap's conception of empirical objects needs to be revised after all.

So far I have treated the empirical part of languages as an undivided unit, but we need to draw some distinctions now. For while tigers and electrons both differ from numbers and sets in not being abstract, and by playing a role in the description of empirical facts, there is nevertheless an important difference: while tigers can be observed with one's eyes, electrons are unobservable entities that are posited to explain observable phenomena. Carnap therefore calls them *theoretical entities*, and accounting for their role in theories is one of his main areas of interest in the 1950s. In section 4.1.3 we already saw that Carnap liberalised his strict verificationism in order to make room for *disposition predicates* through reduction sentences in "Testability and Meaning" (Carnap 1936, Carnap 1937b), and this trend continued, culminating in the influential "The Methodological Character of Theoretical Concepts" from 1956 (Carnap 1956c).

The account of theoretical terms Carnap gives in this paper is complex, and for our purposes it will not be necessary to explain and assess it in full detail. What is important is that Carnap draws a distinction between the *observation language* (called L_O) and the *theoretical language* (L_T). The latter contains both the means to talk about abstract entities and theoretical entities, and the question Carnap tries to answer is, roughly, how talk about theoretical entities can be explained in terms of the observational vocabulary from L_O. One thing that is particularly striking is that Carnap begins by enumerating a number of requirements that, according to some philosophers, languages must fulfil. One of these requirements is the following:

> Requirement of *nominalism*: the values of the variables must be concrete, observable entities (e.g., observable events, things, or thing-moments). (Carnap 1956c: 41)

And while, unsurprisingly, Carnap denies that such a requirement is justified for languages in general, he does say that it makes sense to accept it for the observation language in particular:

> Since L_O is intended for the description of observable events and therefore is meant to be completely interpreted, the requirements, or at least some of them, seem to have merit. [...] Any language fulfilling these requirements is more directly and more completely understandable than languages transgressing these limitations. (Carnap 1956c: 41)

This is intriguing for two reasons. First, it seems that at this point Carnap has left behind the hyper-Pythagorean conception of a coordinate language for the observation language. The second, and more important, potential upshot requires some setting up.

In the passage I quoted Carnap calls the observation language 'completely interpreted', which distinguishes it from the theoretical language that, so Carnap frequently writes, only has a 'partial interpretation'. As Putnam has pointed out in a well-known paper, however, it is not particularly clear what Carnap really means by the distinction between complete and partial interpretation, since he does not give explicit definitions of these notions anywhere (Putnam 1962). The quote above, however, makes it very tempting to spell out this distinction in ontological terms. For Carnap does seem to say that the fact that the observation language has a complete interpretation is a *consequence* of the fact that the variables of the observation language refer to concrete entities. If this were correct, then ontology would play an important explanatory role in Carnap's account after all.

Of course this kind of interpretation only makes sense if Carnap denied that theoretical terms, which he regards as partially interpreted, refer to concrete entities. And this is in fact what he does. The Ramsey-sentence method plays a major role in Carnap's 1950s theory of theoretical terms, through which sentences with *names* for theoretical entities like electrons are replaced by *existential generalisations*.[2] Carl Gustav Hempel had assumed that Carnap's goal was to eliminate ontological commitments to theoretical entities in this way, and criticised his approach by pointing out that, regardless of whether names or quantifiers are used, the ontology remains unchanged. To this Carnap responded as follows:

> I agree with Hempel that the Ramsey-sentence does indeed refer to theoretical entities by the use of abstract variables. However, it should be noted that these entities are not unobservable physical objects like atoms, electrons, etc., but rather [...] purely logico-mathematical entities, e.g., natural numbers, classes of such, classes of classes, etc. (Carnap 1963b: 963)

It is therefore not surprising that the distinction between the completely interpreted observation language and the partially interpreted theoretical language has been interpreted in analogy with the distinction between *realism* and *instrumentalism* in the philosophy of science. Lavers is especially explicit on this: he reads Carnap as shying away from realism about theoretical entities, which would manifest itself in the claim that 'electron' refers to electrons, and opting for sets and numbers as *ersatz referents* instead. Lavers thinks that this approach is philosophically unmotivated, however:

> Carnap's proposed semantic systems are quite strongly revisionist. Instead of taking a term like 'electron' to stand for an unobservable

2 For more background see Psillos 2000a and Psillos 2000b.

> physical object, Carnap proposes to have it stand for a mathematical entity of some sort. [...] It is not at all clear that having 'electron' refer to a class of natural numbers is more clear, fruitful, or simple than having it stand for electrons, especially if at the level of the object language the term electron is taken to be sufficiently clear. The move made by Carnap in his response to Hempel lacks sufficient motivation. Just as Carnap sought to allow empiricists to overcome their nominalistic scruples, he himself should have more fully overcome his instrumentalist scruples. (Lavers 2016a: 217)

I think that one needs to be more careful here, however. Lavers may well be right that, from a practical point of view, the 'realistic' interpretation he favours is equally good, or even superior, to the 'instrumentalist' proposal Carnap prefers. But the talk of 'overcoming instrumentalist scruples' suggests that the decision between these two options is of great philosophical importance, and might even amount to a fundamental revision of Carnap's position, which I think is mistaken. For in the "Methodological Character" paper itself Carnap addresses the question of what theoretical terms refer to, and warns his readers not to overinterpret some of his claims:

> I said previously that the elements of the basic domain I may be regarded as natural numbers. But I warned that this remark and the others about real numbers, etc. should not be taken literally but merely as a didactic help by attaching familiar labels to certain kinds of entities or, to say it in a still more cautious way, to certain kinds of expressions in L_T. [...] Thus the structure can be uniquely specified but the elements of the structure cannot. Not because we are ignorant of their nature; rather, there is no question of their nature. (Carnap 1956c: 45f)

This warning should be taken very seriously. I think that, on closer inspection, the initial impression that the distinction between the completely interpreted observation language and the partially interpreted theoretical language has to do with a difference in their respective ontologies is misguided after all. This is clearer in some of Carnap's earlier writings from the mid-1930s, for there observation terms are not characterised with reference to ontology, but rather by drawing on the ability to *use* them in a certain way:

> A predicate 'P' of a language L is called *observable* for an organism (e.g. a person) N, if, for suitable arguments, e.g. 'b', N is able under

> suitable circumstances to come to a decision with the help of few observations about a full sentence, say 'P(b)', i.e. to a confirmation of either 'P(b)' or '∼P(b)' of such a high degree that he will either accept or reject 'P(b)'. (Carnap 1936: 454f)

That this is the distinguishing feature of observational terms over theoretical terms becomes even more explicit in the following passage:

> We can, of course, state a rule for any term, no matter what its degree of abstractness, in a form like this: 'the term 'te' designates temperature', provided the metalanguage used contains a corresponding expression (here the word 'temperature') to specify the designatum of the term in question. But suppose we have in mind the following purpose for our syntactical and semantical description of the system of physics: the description of the system shall teach a layman to understand it, i.e., to enable him to apply it to his observations in order to arrive at explanations and predictions. [...] A rule like 'the sign 'P' designates the property of being blue' will do for the purpose indicated; but a rule like 'the sign 'Q' designates the property of being electrically charged' will not do. (Carnap 1939: 62)

Here Carnap says that we *could* adopt the approach Lavers suggests, i.e. to give a 'realistic' interpretation of theoretical terms. But he stresses that while this is possible it is not particularly useful, since the important question about theoretical terms is not what they refer to, but rather how they are to be *used*. For observational terms like 'blue', so Carnap supposes, how and when to apply them is already understood, and so in those cases there is no need for anything more informative than disquotational application conditions. Not so for theoretical terms, however, especially if they are newly introduced. Bridge principle of the following kind are therefore crucial:

> If x is warmer than y, then the temperature of x is higher than the temperature of y.

Such principles connect statements involving theoretical terms to already understood statements involving observational vocabulary. Since few theoretical terms can be *explicitly defined* in terms of observational terms, however, there will in all likelihood be cases where the bridge rules do not specify whether a theoretical term applies or not. And as I understand Carnap, *this* is precisely the sense in which the interpretation of theoretical terms is not complete but merely *partial*. While he assumes that for observational terms we can always decide whether they apply in a particular situation or not, in the case of theoretical terms it will sometimes be left indeterminate whether they apply. In fact we already encountered this phenomenon for the special case of disposition predicates in sections 4.1.3 and 7.4.

Whether this is a plausible account of theoretical entities is not something I want to decide here. But if my interpretation is on the right track, questions of reference and ontology are actually quite unimportant in spelling out the nature of the observation language – just as one would have expected, and in spite of the potentially misleading passages I discussed.

9.3 Making Contact with Reality

The upshot of the previous two sections can be summarised as follows: although at first sight abstract objects seem to play an important role in Carnap's account of the empirical world, this initial impression must be severely qualified. On closer inspection objects, whether concrete *or* abstract, do not really play an explanatory role in how language makes contact with the world at all, and hence the thesis of hyper-Pythagoreanism is much less strange and radical than it sounds at first. This observation raises the question whether Beth's argument from non-standard models has *any* relevance to the issues just discussed, or whether our excursus into the empirical world was a mere digression. In the following I will argue that there *are* some aspects of Beth's criticism that generalise beyond the philosophy of mathematics.

Considered at a high level of abstraction, one can draw the following lesson from my interpretation of Beth's argument. In his published works Carnap tends to focus on the description and construction of formal linguistic frameworks, and pays little explicit attention to how these relate to language as it is actually used. This suggests that the latter question is not important for Carnap's project, but Beth's argument has shown that this is mistaken. Some of the challenges for Carnap's position only begin to emerge when we consider the relationship between formal and natural languages, so anyone interested in an assessment of Carnap's position is well-advised to pay special attention to it.

This is a lesson, so I think, that also applies to Carnap's account of the empirical world. I have repeatedly asserted that, for Carnap, language doesn't make contact with reality by an acquaintance relation that connects names with objects. But how *does* hooking up language with reality work then? Suppose we have constructed a formal calculus which is as yet completely uninterpreted. Carnap thinks that, if we want to, we can give such a calculus an empirical interpretation, which in turn enables us to use the formal system to talk about the world. What we need to investigate in more detail, however, is what he means by 'giving an empirical interpretation'.

Logical Syntax contains a section called "The Interpretation of a Language", which addresses precisely this question. Here interpretation is equated with the *translation* of the formal language into a natural language that is assumed to be already understood:

> 'P_1' shall be equivalent in meaning to 'red', ..., 'P_k' to 'blue' [...] (Carnap 1937a: 230)

One might suppose that the later acceptance of semantic notions, especially reference, changed Carnap's conception of interpretation.[3] But the differences between syntactic and semantic period are actually very minor, for while semantic notions are now being used, the important theoretical work is still done by the assumption that the background language in which the interpretation is stated is already understood. Here are two representative examples from Carnap's semantic period:

> The *customary interpretation*, i.e., that for whose sake the calculus is constructed, is given by the following semantical rules. 'lg(x, t)' designates the length in centimeters of the body x at the time t (defined by the statement of a method of measurement) ; 'te(x, t)' designates the absolute temperature in centigrades of x at the time t (likewise defined by a method of measurement); 'th(x)' designates the coefficient of thermic expansion for the body x; 'Sol' designates the class of solid bodies; 'Fe' the class of iron bodies. (Carnap 1939: 58)
>
> For example, the rule
>
> (2) "DDes^e (c_1, Blue)"
>
> says that the direct designatum of the constant c_1 is the class Blue, i.e., the class of those positions which are blue. (Carnap 1963b: 900)

[3] Beth in fact suggests this to be so (Beth 1963: 484).

As such there is of course nothing wrong with this approach. It makes perfect sense to hook up an as-yet meaningless calculus with the real world by means of another language which has already made contact with reality. But things become more problematic once we begin to wonder how the background language we use to achieve this task manages to describe an empirical reality in the first place. It seems that at this point we have two options: either we countenance a way of giving an interpretation that does *not* already rely on an interpreted language, or we regard an interpreted background language as basic, and deny that there are any useful questions to be asked about how this language relates to the world.

The second option seems to be the view Wittgenstein endorses in the *Tractatus*. It is plausible to interpret him as holding that certain fundamental questions about language that appear sensible, including those about how words relate to objects, are actually nonsense. For Wittgenstein there is a sense in which we cannot talk *about* language at all, since its fundamental features cannot be *described* but merely *show themselves*. How exactly to understand this doctrine is very controversial, but it is reasonable to read it as denying the possibility of investigations into how language makes contact with reality.

It is not likely that Carnap would have approved of this Wittgensteinian view, however. In *Logical Syntax*, for instance, he explicitly rejects Wittgenstein's idea of *unsayability*:

> According to another opinion (that of Wittgenstein), there exists only one language, and what we call syntax cannot be expressed at all – it can only "be shown." As opposed to these views, we intend to show that, actually, it is possible to manage with one language only; not, however, by renouncing syntax, but by demonstrating that without the emergence of any contradictions the syntax of this language can be formulated within this language itself. (Carnap 1937a: 53)

And the same sentiment is still present many years later in Carnap's Intellectual Autobiography, from which I already quoted the relevant passage in section 5.2.2. If he does not want to go down the Tractarian route, then, Carnap should adopt the first option: give an account of how language relates to the world that does not rely on the existence of an already understood language.

I am not sure whether such an account *could* be given, but it is clear that Carnap did not in fact have much interest in pursuing a project of this kind. The question of how language makes contact with the world was most salient during the *protocol sentence debate* of the early 1930s.[4] The details are complex, but in his interventions Carnap puts most emphasis on the claim that there is no *one true way* of formulating protocol-sentences, and seems to regard most other substantial questions as matters for biology and psychology to investigate, not philosophers (Carnap 1932). This attitude already seemed unsatisfactory to some of Carnap's contemporaries, such as Ernest Nagel, who writes the following in a report about the state of philosophy in Europe:

> Since Carnap does not believe that science is simply human caprice, and he does think there are better reasons for acting upon predictions made on the basis of scientific method than by a clairvoyant, it is essential that he specify more carefully than he has yet done the procedures involved in arriving at the protocols. Indeed, I think Carnap has thrown out the baby with the bath in excluding from philosophy the study of the process of observation through which protocols are obtained. For while it is legitimate to departmentalize research and to define logic as the study of syntax, a coherent account of meaning and knowledge is sacrificed by not considering the more inclusive situation in which thinking has its origin, development, and terminus. (Nagel 1936: 45f)

Despite such complaints Carnap did not consider the observational language to be of great philosophical importance in his later work. In the writings on theoretical entities we briefly looked at, for instance, Carnap regards the observation language as unproblematic and fully understood, and thus quickly proceeds to talk about the issues he regards as less clear. One will therefore not find a worked-out account of the language-world relation in Carnap's writings.

I think that this is an important gap. Whether there is a *problem* cannot be decided here, however. Some of Carnap's contemporaries, in particular Neurath and Quine, were much more interested in this particular issue after all, and it might well be that their views are compatible with Carnap's outlook. This question goes beyond what I can investigate here. I hope to have established, however, that investigating the relationship between formal frameworks and natural language is important not only for Carnap's philosophy of mathematics, but also for his theory of the empirical world.

[4] For all the twists and turns of this debate see Uebel 2007.

9.4 Thomasson's Easy Ontology

I will end this chapter by briefly considering a contemporary position that can be described *neo-Carnapian*: Amie Thomasson's easy ontology. As before, my ambition is not to give a comprehensive summary and evaluation of this position. I rather want to show that this modern view faces questions about how contact with the world is made as well, and that Thomasson is not more explicit about these issues than Carnap was.

Thomasson's easy approach to ontology is opposed to what has become the mainstream methodology for doing metaphysics, namely the neo-Quinean view according to which we can find out what exists by looking at what the quantifiers of the best overall theory of the world range over (Thomasson 2007, Thomasson 2009, Thomasson 2015). According to Thomasson, the construction and comparison of sophisticated metaphysical theories that is characteristic of this methodology is not actually needed. Instead we can answer ontological questions easily using simple empirical and conceptual means. Here the concept of an *application condition* plays a major role: Thomasson thinks that sortal terms such as 'table' and 'chair' are associated with conditions of the following form:

(T) 'Table' refers iff there are particles arranged table-wise.

Furthermore, she endorses the following link between reference and existence:

(E) Ks exist iff the application conditions actually associated with 'K' are fulfilled. (Thomasson 2015: 86)

And since Thomasson regards it as an obvious empirical truth that there are particles arranged table-wise – one that is not even challenged by revisionary metaphysicians who deny the existence of composite objects – one can establish that ordinary objects such as tables exist without doing any deep metaphysics.

A lot more could be said about the account than I did in this very brief sketch. Since this is principally a thesis about Carnap, I only want to comment on the extent to which Thomasson's position resembles that of Carnap. Thomasson is keen to stress the Carnapian origins of her positions, and it is certainly true that both of them share a deflationary attitude towards ontology and metaphysics. It must be noted, however, that some aspects of Thomasson's position are also very un-Carnapian. As we have seen for Carnap the distinction between linguistic frameworks with explicit rules, and the informal natural language which does not uniquely pin down a particular set of rules, is crucial. This distinction makes the principle of tolerance possible and gives rise to the need for explications. Thomasson, on the other hand, seems to think that the linguistic rules which she calls application conditions can be immediately read off natural language, and that hence our ordinary concepts commit us to a certain ontology. In response to ontological questions such as whether there are ordinary objects, her answer is therefore not 'it depends on what linguistic framework for talk about ordinary objects we want to accept', but rather the uncompromising 'yes'.

These un-Carnapian tendencies are especially apparent in her philosophy of mathematics. Although she doesn't discuss her account in that much detail, it does not very much resemble Carnap's *Logical Syntax*, but rather seems to be a form of *neo-Fregean* logicism, with Thomasson responding to classic problems such as the bad company objection (Thomasson 2015: chapter 8). Since introducing neo-Fregeanism would take us too far afield I will therefore set Thomasson's philosophy of mathematics aside. Instead I want to mention one aspect of her view that immediately connects with the topic of the previous section: namely the role of application conditions in an account of how language makes contact with reality.

The relevant point is best introduced by considering an objection against easy ontology recently posed by Andrew Brenner. One principle about application condition Thomasson endorses is the following:

(¬C) Application conditions must not take the following form: 'K' applies iff Ks exist. (While this will always be true, it will not count as an application condition, in our terms.) (Thomasson 2015: 96)

Based on this, Brenner notes that Thomasson's view seems to face either an infinite regress or a problematic kind of circularity. For suppose A is a sortal with non-trivial application conditions, which means that whether 'A' refers or not cannot be settled on the basis of conceptual truths alone, but requires some worldly condition to be fulfilled. Due to (¬C), the application conditions for A cannot be "A' refers iff As exist', but rather will have to be something like "A' refers iff Bs exist', for some sortal B. However, whether Bs exist depends on whether the application conditions associated with 'B' are fulfilled. Since those cannot be "B' refers iff Bs exist', another sortal C is needed: "B' refers iff Cs exist'. It is easily seen that this reasoning can be repeated indefinitely, and so there either is an infinite regress, or there must be exceptions to (¬C) after all. And as Brenner points out, neither option seems acceptable for Thomasson (Brenner 2018: 4f).

This argument is watertight if we assume that, necessarily, on the right-hand side of an application condition for a sortal a different sortal must be invoked. Things are different, however, if there could be application conditions of the form

'A' refers iff Φ,

where Φ expresses a non-trivial condition but doesn't do so by means of a sortal. In that case the regress could terminate with application conditions that are *wordly* – i.e. require that the world is a certain way rather than another to be fulfilled – but do not themselves depend on yet further application conditions.

Whether this way out of the regress is an option for Thomasson depends on what the precise role of sortals and application conditions is supposed to be on her view. At times it is tempting to read her as saying that sortals with application conditions are needed for language to make contact with the world *at all*. If that is the right reading, then Brenner's argument is certainly forceful. But other passages suggest that her actual view is more subtle: while she certainly thinks that sortals are required to represent the world in terms of persistent *objects* with determinate identity-conditions across space and time, Thomasson also sees room for pre-objectual ways to refer to reality (Thomasson 2015: section 2.3).

Brenner recognises that this is a possible way out:

> Perhaps Thomasson can avoid the sort of regress or circularity I'm pressing by maintaining that there can be sortal terms with non-trivial application conditions which nevertheless are such that those application conditions do not appeal to the existence of anything for their satisfaction. Thomasson makes at least two suggestions that might be amenable to a view of this sort. In both cases, Thomasson aims to specify sufficient application conditions for there being objects which fall under some sortal K which are such that the sufficient application conditions in question do not appeal to there being objects of any sort [...] (Brenner 2018: 4)

The first of the two suggestions Brenner alludes to here is to give sortal-free application conditions by relying on the idea that there is a basic concept of an object that can be explained purely in terms of perceptual experience. Drawing on research in contemporary psychology, Thomasson argues that there is a notion of *object of experience* that even infants can use, and whose application conditions won't involve any sortal. The second approach to give sortal-free application conditions is to use a feature-placing language (O'Leary-Hawthorne and Cortens 1995). Using this, we can for instance state the application conditions for *cup* as follows:

(C) 'Cup' refers iff it is cupping around here.

Thomasson's discussions of these two options are quite brief however, and Brenner puts forward objections to both approaches. I do not want to assess whether these are successful here, however. Even if Thomasson can be defended, Brenner has made an important point: namely that the question of how language makes contact with reality in the first place is not something a defender of easy ontology can just set aside, but is rather central to the stability of the position. Unfortunately – just like Carnap – Thomasson spends comparatively little time on this topic in her writings.

In fact, Thomasson is not the only neo-Carnapian who has been criticised for being insufficiently clear about the fundamental question of the language-world relation. Rayo's position in *The Construction of Logical Space*, for instance, resembles Thomasson's position in that conditionals stating truth-conditions play a major explanatory role (Rayo 2013). As Jason Turner points out, however, this cannot be the whole story about how language makes contact with the world, since the specifications Rayo (and also Thomasson) describe rely on other parts of language:

> Rayo talks quite a lot about 'truth-conditions,' and assigning truth-conditions to sentences. But nowhere in the book do we get a *metaphysics* of truth-conditions. We're told how to 'specify' them, but the specification is of course always simply in terms of other sentences, which in turn need truth-conditions specified for them, and so on down the line. There's no story as to how the linguistic rubber meets the extra-linguistic road, to get the whole enterprise running down the track. (Turner 2015: 2614)

This is a slightly unfortunate way of putting the problem, since a *metaphysics* of truth-conditions sounds like the kind of thing Carnap and the neo-Carnapians would want to reject in any case. But, using another metaphor, the point can be put in such a way that it cannot be ignored as easily: what seems needed is some story about why our language (in the case of Thomasson) or a linguistic framework (in the case of Carnap) is not just "frictionless spinning in the void"

(McDowell 1994: 11) – i.e. why it is that ”there's a tiger in a room” is true in some situations and not others, depending on factors external to language itself.

The gap in Carnap's account of how language makes contact with the empirical world thus reappears. It seems to me that neo-Carnapians like Thomasson have two ways to respond: either they could supplement their account by a positive story, or they could try to *dissolve* the problem by denying the assumption that such a story is needed in the first place. I suspect that Thomasson would prefer the second route, but will not be able to go into this topic here. For now it is safe to conclude that engaging with Carnap on this matter is of more than historical interest.

Conclusion

In this thesis I have argued that Carnap's anti-metaphysical philosophy of mathematics cannot be accepted as it stands. I will conclude by summarising the possible reactions to this result, assuming that my argument is correct. Some of these options have already been discussed earlier on, but a concise overview should be useful for friends and enemies of Carnap's position alike. During the course of this thesis our attention has been focused on the problematic aspects of Carnap's position, but the range of available theoretical options hopefully shows that such a critical engagement can give rise to positive impulses for the philosophy of mathematics.

As I have emphasised several times, my interpretation Carnap's philosophy of mathematics is contentious. I think, however, that there are good exegetical and systematic reasons to read him as accepting the following position: there is a definition of 'analytic' which is such that (1) every mathematical sentence, including undecidable ones such as Con_{PA}, is either determinately analytic or contradictory, and (2) our linguistic behaviour commits us to the acceptance of this definition. In section 6.2 I then argued that giving up (2) would turn Carnap's position into Wittgenstein's radical conventionalism, a view that is universally rejected.

It might of course be that the Wittgensteinian path is more defensible than has been realised. I am not aware of any contemporary philosopher who explicitly endorses a version of radical conventionalism, but what comes closest is probably the idea that mathematical indeterminacy is much more prevalent than is commonly assumed. Joel Hamkins, for instance, has recently suggested that the possibility of non-standard models should undermine our confidence in the claim that there is such a thing as *the* standard model of arithmetic:

> So why are mathematicians so confident that there is an absolute concept of finite natural number, independent of any set-theoretic concerns, when all of our categoricity arguments are explicitly set-theoretic and require one to commit to a background concept of set?

> My long-term expectation is that technical developments will eventually arise that provide a forcing analogue for arithmetic, [...] 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)[5]

This would have wide-ranging consequences indeed, since without a fixed conception of the natural numbers other syntactic notions, such as that of a finite proof, become indeterminate as well. If Hamkin's view is defensible, then the arguments from syntax in the previous chapters would loose some of their bite. I cannot go further into the viability of this option here, but it should be noted that Hamkin's account seems to be a revisionary one, and hence does not accord with Carnap's ambition to leave classical mathematics as it is.

Suppose we do not want to follow Wittgenstein and Hamkins, and stick to the assumption that claims about syntax have determinate truth values. One way to implement this strategy is *exceptionalism* about arithmetic: the idea being that while the adoption of other mathematical theories is a matter of pragmatic decision, arithmetic tracks some genuine features of reality. This is the option Beth himself seems to have in mind, and, as we saw in section 7.1, can also be found in the writings of Hofweber and Clarke-Doane. This represents a departure from Carnap's global internalism, and so to a certain extent the position of the external Platonist is vindicated. But, so the hope would be, internalism about other areas can be saved, and the philosophical questions raised by an ontology of natural numbers, or facts about them, might be more tractable than they would be for other kinds of abstract objects.

Nevertheless, it would be much more exciting if Carnap's own ambitious form of internalism could be salvaged after all, and there is indeed a glimmer of hope. As described in section 7.4, recent work by Warren suggests that there is a way to conceive of infinitary rules on which following them does not require mysterious non-recursive abilities. If this view is defensible, then using the ω-rule might be a way to pin down the standard model without relying on ontology after all. I argued that it is unlikely that this would enable one to retain Carnap's original position without *any* revisions, since Warren's account seems to require a more robustly realist account of dispositions than Carnap has to offer. This kind of departure may be less problematic than other approaches,

[5] See `https://cs.nyu.edu/pipermail/fom/2017-November/020689.html` for a critical discussion of this idea.

however, since an ontology of dispositions seems more tractable by scientific means than an ontology of abstract objects or mathematical facts.

Lastly, we could try to follow Quine and question Carnap's basic assumption that, in order to be compatible with empiricism, mathematics needs to be construed as analytic. In chapter 8 I already flagged a number of potential problems for Quine's philosophy of mathematics, but I think that its viability should be investigated further. Despite seeming quite different on the surface, Quine's philosophy actually preserves many of Carnap's most central commitments, such as his empiricism and the rejection of metaphysical speculation. In light of the problem Beth poses for Carnap's own internalism, those who still want to avoid metaphysics in the philosophy of mathematics might thus be well-advised to turn to Quine.

Bibliography

Alspector-Kelly, M. (2001): 'On Quine on Carnap on Ontology'. *Philosophical Studies*, 102(1):93–122.

Arnold, J. and Shapiro, S. (2007): 'Where in the (World Wide) Web of Belief Is the Law of Non-Contradiction?' *Noûs*, 41(2):276–297.

Awodey, S. and Carus, A. W. (2004): 'How Carnap Could Have Replied to Gödel'. In: Awodey, S. and Klein, C. (eds), *Carnap Brought Home: The View from Jena*, pp. 203–223. Open Court, Chicago and LaSalle, Illinois.

Awodey, S. and Carus, A. W. (2010): 'Gödel and Carnap'. In: Gödel, K., Feferman, S., Parsons, C., and Simpson, S. G. (eds), *Kurt Gödel: Essays for His Centennial*. Cambridge University Press.

Ayer, A. J. (1936): *Language, Truth, and Logic*. London: V. Gollancz.

Bays, T. (2014): 'Skolem's Paradox'. In: Zalta, E. N. (ed), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2014.

Benacerraf, P. (1985): 'Skolem and the Skeptic'. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 59:85–115.

Benacerraf, P. and Putnam, H. (eds) (1983): *Philosophy of Mathematics: Selected Readings*. Cambridge University Press.

Bennett, K. (2009): 'Composition, Colocation, and Metaontology'. In: Chalmers, D. J., Manley, D., and Wasserman, R. (eds), *Metametaphysics: New Essays on the Foundations of Ontology*, pp. 38–76. Oxford University Press.

Bergström, L. (1994): 'Quine's Truth'. *Inquiry*, 37(4):421–435.

Bernays, P. (1998): 'Reply to the Note by Mr. Aloys Müller, "On Numbers as Signs"'. In: Mancosu, P. (ed), *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s*, pp. 223–226. Oxford University Press.

Beth, E. W. (1963): 'Carnap's Views on the Advantages of Constructed Systems over Natural Languages in the Philosophy of Science'. In: Schilpp, P. (ed), *The Philosophy of Rudolf Carnap*, pp. 469–502. Open Court, La Salle, Illinois.

Boghossian, P. (1989): 'The Rule-Following Considerations'. *Mind*, 98(392):507–49.

Boghossian, P. A. (2017): 'Analyticity'. In: *A Companion to the Philosophy of Language*, pp. 578–618. John Wiley & Sons, Ltd.

Bohnert, H. G. (1975): 'Carnap's Logicism'. In: Hintikka, J. (ed), *Rudolf Carnap, Logical Empiricist: Materials and Perspectives*, pp. 183–216. Reidel, Dordrecht.

Boolos, G. (1984): 'To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables)'. *Journal of Philosophy*, 81(8):430–449.

Bradley, D. (2018): 'Carnap's Epistemological Critique of Metaphysics'. *Synthese*, 195(5):2247–2265.

Brenner, A. (2018): 'Easy Ontology, Application Conditions and Infinite Regress'. *Analysis*, 78(4):605–614.

Bronzo, S. (2017): 'Wittgenstein, Theories of Meaning, and Linguistic Disjunctivism'. *European Journal of Philosophy*, 25(4):1340–1363.

Broughton, G. L. (forthcoming): 'Carnapian Frameworks'. *Synthese*.

Buldt, B. (2004): 'On RC 102-43-14'. In: Awodey, S. and Klein, C. (eds), *Carnap Brought Home: The View from Jena*, pp. 225–246. Open Court, Chicago and LaSalle, Illinois.

Buldt, B. (2014): 'The Scope of Gödel's First Incompleteness Theorem'. *Logica Universalis*, 8(3-4):499–552.

Button, T. (2013): *The Limits of Realism*. Oxford University Press.

Button, T. and Walsh, S. (2018): *Philosophy and Model Theory*. Oxford University Press.

Carnap, R. (1928): *Scheinprobleme in der Philosophie: Das Fremdpsychische und der Realismusstreit*. Weltkreis Verlag, Berlin.

Carnap, R. (1931): 'Überwindung der Metaphysik durch logische Analyse der Sprache'. *Erkenntnis*, 2:219–241.

Carnap, R. (1932): 'Über Protokollsätze'. *Erkenntnis*, 3(1):215–228.

Carnap, R. (1934): *The Unity of Science*. Translated by Max Black. K. Paul, Trench, Trubner & Co., London.

Carnap, R. (1936): 'Testability and Meaning'. *Philosophy of Science*, 3(4):419–471.

Carnap, R. (1937a): *The Logical Syntax of Language*. K. Paul, Trench, Trubner & Co., London.

Carnap, R. (1937b): 'Testability and Meaning–Continued'. *Philosophy of Science*, 4(1):1–40.

Carnap, R. (1938): 'Review: Barkley Rosser, Gödel Theorems for Non-Constructive Logics'. *Journal of Symbolic Logic*, 3(1):50–50.

Carnap, R. (1939): 'Foundations of Logic and Mathematics'. In: Neurath, O., Carnap, R., and Morris, C. (eds), *International Encyclopedia of Unified Science*, pp. 139–213. University of Chicago Press.

Carnap, R. (1942): *Introduction to Semantics*. Harvard University Press.

Carnap, R. (1950): *Logical Foundations of Probability*. University of Chicago Press.

Carnap, R. (1952): 'Meaning Postulates'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 3(5):65–73.

Carnap, R. (1953): 'Formal and Factual Science'. In: Feigl, H. and Brodbeck, M. (eds), *Readings in the Philosophy of Science*, pp. 123–128. Appleton-Century-Crofts, New York.

Carnap, R. (1955): 'Meaning and Synonymy in Natural Languages'. *Philosophical Studies*, 6(3):33–47.

Carnap, R. (1956a): 'Empiricism, Semantics, and Ontology'. In: *Meaning and Necessity*, pp. 205–221. University of Chicago Press.

Carnap, R. (1956b): *Meaning and Necessity*. University of Chicago Press.

Carnap, R. (1956c): 'The Methodological Character of Theoretical Concepts'. *Minnesota Studies in the Philosophy of Science*, 1(1):38–76.

Carnap, R. (1958): *Introduction to Symbolic Logic and Its Applications*. Dover Publications, New York.

Carnap, R. (1963a): 'Intellectual Autobiography'. In: Schilpp, P. A. (ed), *The Philosophy of Rudolf Carnap*, pp. 1–84. Open Court, La Salle, Illinois.

Carnap, R. (1963b): 'Replies and Systematic Expositions'. In: Schilpp, P. A. (ed), *The Philosophy of Rudolf Carnap*, pp. 859–1013. Open Court, La Salle, Illinois.

Carnap, R. (1967): *The Logical Structure of the World*. University of California Press, Berkeley.

Carnap, R. (1968): 'Inductive Logic and Inductive Intuition'. In: Lakatos, I. (ed), *The Problem of Inductive Logic*, pp. 258–314. North Holland Pub. Co., Amsterdam.

Carnap, R. (1983): 'The Logicist Foundations of Mathematics'. In: Benacerraf, P. and Putnam, H. (eds), *Philosophy of Mathematics: Selected Readings*, pp. 41–52. Cambridge University Press.

Carnap, R. (1990): 'Quine on Analyticity'. In: Creath, R. (ed), *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work: Edited and with an Introduction by Richard Creath*, pp. 428–433. University of California Press, Berkeley.

Carnap, R. (forthcoming): *Tagebücher 1908–1935.* Edited by Christian Damböck. Meiner, Hamburg.

Carus, A. W. (2004): 'Sellars, Carnap, and the Logical Space of Reasons'. In: Awodey, S. and Klein, C. (eds), *Carnap Brought Home: The View from Jena*, pp. 317–355. Open Court, Chicago and LaSalle, Illinois.

Carus, A. W. (2007): *Carnap and Twentieth-Century Thought: Explication as Enlightenment*. Cambridge University Press.

Carus, A. W. (2017): 'Carnapian Rationality'. *Synthese*, 194(1):163–184.

Cat, J. and Tuboly, A. T. (eds) (2019): *Neurath Reconsidered: New Sources and Perspectives*. Springer.

Clarke-Doane, J. (2020): *Morality and Mathematics*. Oxford University Press.

Coffa, A. (1987): 'Carnap, Tarski and the Search for Truth'. *Noûs*, 21(4):547–572.

Coffa, A. (1991): *The Semantic Tradition From Kant to Carnap: To the Vienna Station*. Cambridge University Press.

Colyvan, M. (2012): *An Introduction to the Philosophy of Mathematics*. Cambridge University Press.

Copi, I. M. (1949): 'Modern Logic and the Synthetic a Priori'. *Journal of Philosophy*, 46(8):243–245.

Copi, I. M. (1950): 'Gödel and the Synthetic a Priori: A Rejoinder'. *Journal of Philosophy*, 47(22):633–636.

Creath, R. (1990): *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work: Edited and with an Introduction by Richard Creath*. University of California Press, Berkeley.

Creath, R. (1991): 'Every Dogma Has Its Day'. *Erkenntnis*, 35(1):347–389.

Creath, R. (2003): 'The Linguistic Doctrine and Conventionality: The Main Argument in "Carnap and Logical Truth"'. In: Hardcastle, G. L. and Richardson, A. W. (eds), *Logical Empiricism in North America*, pp. 234–256. University of Minnesota Press, Minneapolis.

Creath, R. (2007): 'Quine's Challenge to Carnap'. In: Friedman, M. and Creath, R. (eds), *The Cambridge Companion to Carnap*, pp. 316–335. Cambridge University Press.

Detlefsen, M. (1986): *Hilbert's Program: An Essay on Mathematical Instrumentalism*. Reidel, Dordrecht.

Diamond, C. (1991): 'The Face of Necessity'. In: *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*, pp. 243–266. The MIT Press.

Doyle, E. (2013): *The Methodological Roles of Tolerance and Conventionalism in the Philosophy of Mathematics: Reconsidering Carnap's Logic of Science*. PhD thesis, The University of Western Ontario.

Dreben, B. (1996): 'Quine and Wittgenstein: The Odd Couple'. In: Arrington, R. L. and Glock, H.-J. (eds), *Wittgenstein and Quine*, pp. 39–61. Routledge, London.

Dreben, B. (2006): 'Quine on Quine'. In: Gibson, R. F. (ed), *The Cambridge Companion to Quine*, pp. 287–293. Cambridge University Press.

Dreben, B. and van Heijenoort, J. (1986): 'Introductory Note to *1929*, *1930* and *1930a*'. In: *Kurt Gödel Collected Works. Volume I. Publications 1929-1936*, pp. 44–59. Oxford University Press.

Dummett, M. (1959): 'Wittgenstein's Philosophy of Mathematics'. *Philosophical Review*, 68(3):324–348.

Dummett, M. (1978): *Truth and Other Enigmas*. Harvard University Press.

Ebbs, G. (1997): *Rule-Following and Realism*. Harvard University Press.

Ebbs, G. (2011): 'Carnap and Quine on Truth by Convention'. *Mind*, 120(478):193–237.

Ebbs, G. (2017a): 'Carnap on Ontology'. In: *Carnap, Quine, and Putnam on Methods of Inquiry*, pp. 13–32. Cambridge University Press.

Ebbs, G. (2017b): 'Carnap's Logical Syntax'. In: *Carnap, Quine, and Putnam on Methods of Inquiry*, pp. 33–54. Cambridge University Press.

Ebbs, G. (2019): 'Carnap on Analyticity and Existence: A Clarification, Defense, and Development of Quine's Reading of Carnap's Views on Ontology'. *Journal for the History of Analytical Philosophy*, 7(5):1–31.

Eklund, M. (2007): 'The Ontological Significance of Inscrutability'. *Philosophical Topics*, 35(1-2):115–134.

Eklund, M. (2012): 'Multitude, Tolerance and Language-Transcendence'. *Synthese*, 187(3):833–847.

Eklund, M. (2013): 'Carnap's Metaontology'. *Noûs*, 47(2):229–249.

Fara, R. (ed) (1994): *In Conversation: W.V. Quine.* Series of Video Tapes. Philosophy International, Centre for the Philosophy of the Natural and Social Sciences, London School of Economics.

Feferman, S. (1991): 'Reflecting on Incompleteness'. *Journal of Symbolic Logic*, 56(1):1–49.

Field, H. (1980): *Science Without Numbers*. Princeton University Press.

Field, H. (1988): 'Realism, Mathematics and Modality'. *Philosophical Topics*, 16(1):57–107.

Field, H. (1994): 'Are Our Logical and Mathematical Concepts Highly Indeterminate?' *Midwest Studies in Philosophy*, 19(1):391–429.

Flocke, V. (2019): 'Carnap's Defense of Impredicative Definitions'. *The Review of Symbolic Logic*, 12(2):372–404.

Flocke, V. (2020): 'Carnap's Noncognitivism About Ontology'. *Noûs*, 54(3):527–548.

Franzén, T. (2004): 'Transfinite Progressions: A Second Look at Completeness'. *The Bulletin of Symbolic Logic*, 10(3):367–389.

Frege, G. (1884): *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Wilhelm Koebner, Breslau.

Frege, G. (1893): *Grundgesetze der Arithmetik: Begriffsschiftlich abgeleitet*. Band 1. H. Pohle, Jena.

Frege, G. (1918): 'Der Gedanke'. *Beiträge Zur Philosophie des Deutschen Idealismus*, 2:58–77.

Frege, G. (1979): *Posthumous Writings*. Translated by Peter Long and Roger White. Blackwell, Oxford.

Frege, G. (1980): *Philosophical and Mathematical Correspondence*. Blackwell, Oxford.

Frege, G. (2013): *Basic Laws of Arithmetic*. Translated by Philip A. Ebert and Marcus Rossberg. Oxford University Press.

Friedman, M. (1999a): 'Analytic Truth in Carnap's *Logical Syntax of Language*'. In: *Reconsidering Logical Positivism*. Cambridge University Press.

Friedman, M. (1999b): 'Tolerance and Analyticity in Carnap's Philosophy of Mathematics'. In: *Reconsidering Logical Positivism*, pp. 198–233. Cambridge University Press.

Friedman, M. (2009): 'Tolerance, Intuition, and Empiricism'. In: Wagner, P. (ed), *Carnap's Logical Syntax of Language*, pp. 236–249. Palgrave-Macmillan.

Friedman-Biglin, N. (2015): *Carnap's Conventionalism: Logic, Science, and Tolerance*. PhD thesis, University of St Andrews.

Frost-Arnold, G. (2013): *Carnap, Tarski, and Quine at Harvard: Conversations on Logic, Mathematics, and Science*. Open Court Press, Chicago.

George, A. (2000): 'On Washing the Fur Without Wetting It: Quine, Carnap, and Analyticity'. *Mind*, 109(433):1–24.

Gödel, K. (1983): 'Russell's Mathematical Logic'. In: Benacerraf, P. and Putnam, H. (eds), *Philosophy of Mathematics: Selected Readings*, pp. 447–469.

Gödel, K. (1986): 'Zur intuitionistischen Arithmetik und Zahlentheorie'. In: *Collected Works. Volume I. Publications 1929-1936*, pp. 286–294. Oxford University Press.

Gödel, K. (1995a): 'Is Mathematics Syntax of Language?' In: *Collected Works. Volume III. Unpublished Essays and Lectures.*, pp. 334–355. Oxford University Press: Oxford.

Gödel, K. (1995b): 'Is Mathematics Syntax of Language?' In: Rodriguez-Consuegra, F. (ed), *Kurt Gödel: Unpublished Philosophical Essays*, pp. 171–222. Birkhäuser Basel.

Gödel, K. (2003a): *Collected Works. Volume IV. Correspondence A-G*. Oxford University Press.

Gödel, K. (2003b): *Collected Works. Volume V. Correspondence H-Z*. Oxford University Press.

Goldfarb, W. (1979): 'Logic in the Twenties: The Nature of the Quantifier'. *Journal of Symbolic Logic*, 44(3):351–368.

Goldfarb, W. (1995): 'Introductory Note to *1953/9'. In: *Kurt Gödel Collected Works. Volume III. Unpublished Essays and Lectures.*, pp. 324–333. Oxford University Press.

Goldfarb, W. (1997): 'Semantics in Carnap: A Rejoinder to Alberto Coffa'. *Philosophical Topics*, 25(2):51–66.

Goldfarb, W. (2003): 'Rudolf Carnap'. In: *Kurt Gödel Collected Works. Volume IV. Correspondence A-G*, pp. 335–42. Oxford University Press.

Goldfarb, W. (2005): 'On Gödel's Way In: The Influence of Rudolf Carnap'. *Bulletin of Symbolic Logic*, 11(2):185–193.

Goldfarb, W. (2009): 'Carnap's Syntax Programme and the Philosophy of Mathematics'. In: Wagner, P. (ed), *Carnap's Logical Syntax of Language*. Palgrave-Macmillan.

Goldfarb, W. and Ricketts, T. (1992): 'Carnap and the Philosophy of Mathematics'. In: Bell, D. and Vossenkuhl, W. (eds), *Wissenschaft und Subjektivität. Science and Subjectivity*, pp. 61–78. Akademie Verlag, Berlin.

Goodman, N. and Quine, W. V. (1947): 'Steps Toward a Constructive Nominalism'. *Journal of Symbolic Logic*, 12(4):105–122.

Gregory, P. A. (2003a): 'Putting the Bite Back Into 'Two Dogmas''. *Principia*, 7(1-2):115–129.

Gregory, P. A. (2003b): ''Two Dogmas'–All Bark and No Bite? Carnap and Quine on Analyticity'. *Philosophy and Phenomenological Research*, 67(3):633–648.

Haack, S. (1976): 'Some Preliminaries to Ontology'. *Journal of Philosophical Logic*, 5(4):457–474.

Hale, B. and Wright, C. (2003): *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford University Press.

Hamkins, J. D. (2012): 'The Set-Theoretic Multiverse'. *Review of Symbolic Logic*, 5(3):416–449.

Hart, W. D. (1979): 'The Epistemology of Abstract Objects'. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 53:153–165.

Heck, R. (2007): 'Frege and Semantics'. *Grazer Philosophische Studien*, 75(1):27–63.

Hofweber, T. (2014): 'Rayo's *The Construction of Logical Space*'. *Inquiry*, 57(4):442–454.

Hofweber, T. (2016): *Ontology and the Ambitions of Metaphysics*. Oxford University Press.

Hylton, P. (2006): 'Quine on Reference and Ontology'. In: Gibson, R. F. (ed), *The Cambridge Companion to Quine*, pp. 115–50. Cambridge University Press.

Isaacson, D. (1992): 'Carnap, Quine, and Logical Truth'. In: Vossenkuhl, W. (ed), *Wissenschaft und Subjektivität. Science and Subjectivity*, pp. 360–391. Akademie Verlag, Berlin.

Isaacson, D. (2004): 'Quine and Logical Positivism'. In: Gibson, R. F. (ed), *The Cambridge Companion to Quine*, pp. 214–269. Cambridge University Press.

Kaye, R. (1991): *Models of Peano Arithmetic*. Clarendon Press, Oxford.

Kleene, S. C. (1939): 'Review: Rudolf Carnap, The Logical Syntax of Language'. *Journal of Symbolic Logic*, 4(2):82–87.

Koellner, P. (ms): 'Carnap on the Foundations of Logic and Mathematics'. Unpublished manuscript. Online: http://logic.harvard.edu/koellner/CFLM.pdf.

Köhler, E. (2002): 'Rudolf Carnap – Kurt Gödel. Gespräche und Briefe 1928-1940'. In: Buldt, B., Köhler, E., Weibel, P., Stöltzner, M., and DePauli-Schimanovich-Göttig, W. (eds), *Wahrheit und Beweisbarkeit. Volume 1: Dokumente und historische Analysen*, pp. 109–128. oebv and hpt, Wien.

Kraut, R. (2020): 'Rudolf Carnap: Pragmatist and Expressivist about Ontology.' In: Bliss, R. and Miller, J. T. M. (eds), *The Routledge Handbook of Metametaphysics*, pp. 32–48. Routledge, London.

Kripke, S. A. (1982): *Wittgenstein on Rules and Private Language*. Harvard University Press.

Lavers, G. (2004): 'Carnap, Semantics and Ontology'. *Erkenntnis*, 60(3):295–316.

Lavers, G. (2008): 'Carnap, Formalism, and Informal Rigour'. *Philosophia Mathematica*, 16(1):4–24.

Lavers, G. (2015): 'Carnap, Quine, Quantification and Ontology'. In: Torza, A. (ed), *Quantifiers, Quantifiers, and Quantifiers: Themes in Logic, Metaphysics, and Language*. Springer.

Lavers, G. (2016a): 'Carnap on Abstract and Theoretical Entities'. In: Blatti, S. and Lapointe, S. (eds), *Ontology After Carnap*. Oxford University Press.

Lavers, G. (2016b): 'Carnap's Surprising Views on the Axiom of Infinity'. *Metascience*, 25(1):37–41.

Lavers, G. (2019): 'Hitting a Moving Target: Gödel, Carnap, and Mathematics as Logical Syntax'. *Philosophia Mathematica*.

Leitgeb, H. and Carus, A. (2020): 'Rudolf Carnap'. In: Zalta, E. N. (ed), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2020.

Leng, M. (2007): 'What's There to Know? A Fictionalist Approach to Mathematical Knowledge'. In: Leng, M., Paseau, A., and Potter, M. (eds), *Mathematical Knowledge*. Oxford University Press.

Lewy, C. (1940): 'Logical Necessity'. *Philosophical Review*, 49(1):62–68.

Linnebo, Ø. (2018): 'Platonism in the Philosophy of Mathematics'. In: Zalta, E. N. (ed), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018.

Linnebo, Ø. (2020a): *Philosophy of Mathematics*. Princeton University Press.

Linnebo, Ø. (2020b): 'Predicative and Impredicative Definitions'. *Internet Encyclopedia of Philosophy*.

Maddy, P. (1997): *Naturalism in Mathematics*. Oxford University Press.

McDowell, J. (1994): *Mind and World*. Harvard University Press.

McGee, V. (1991): 'We Turing Machines Aren't Expected-Utility Maximizers (Even Ideally)'. *Philosophical Studies*, 64(1):115–123.

Mellor, D. H. (1974): 'In Defense of Dispositions'. *The Philosophical Review*, 83(2):157–181.

Nagel, E. (1936): 'Impressions and Appraisals of Analytic Philosophy in Europe. II'. *Journal of Philosophy*, 33(2):29–53.

Neurath, O. (1931): 'Soziologie im Physikalismus'. *Erkenntnis*, 2(1):393–431.

O'Grady, P. (1999): 'Carnap and Two Dogmas of Empiricism'. *Philosophical and Phenomenological Research*, 59(4):1015–1027.

O'Leary-Hawthorne, J. and Cortens, A. (1995): 'Towards Ontological Nihilism'. *Philosophical Studies*, 79(2):143–165.

Oliver, A. and Smiley, T. (2016): *Plural Logic*. Second Edition, Revised and Enlarged. Oxford University Press.

Parsons, C. (1986): 'Quine on the Philosophy of Mathematics'. In: Hahn, L. E. and Schilpp, P. A. (eds), *The Philosophy of W.V. Quine*, pp. 369–395. Open Court, Chicago.

Parsons, C. (2007): *Mathematical Thought and Its Objects*. Cambridge University Press.

Parsons, T. (1970): 'Various Extensional Notions of Ontological Commitment'. *Philosophical Studies*, 21(5):65–74.

Pears, D. F. (1987): *The False Prison: A Study of the Development of Wittgenstein's Philosophy*. Oxford University Press.

Potter, M. (2000): *Reason's Nearest Kin: Philosophies of Arithmetic From Kant to Carnap*. Oxford University Press.

Potter, M. (2001): 'Was Gödel a Gödelian Platonist?' *Philosophia Mathematica*, 9(3):331–346.

Potter, M. (2004): *Set Theory and Its Philosophy: A Critical Introduction*. Oxford University Press.

Psillos, S. (2000a): 'Carnap, the Ramsey-Sentence and Realistic Empiricism'. *Erkenntnis*, 52(2):253–279.

Psillos, S. (2000b): 'Rudolf Carnap's 'Theoretical Concepts In Science''. *Studies in History and Philosophy of Science Part A*, 31(4):151–172.

Putnam, H. (1962): 'What Theories Are Not'. In: Nagel, E., Suppes, P., and Tarski, A. (eds), *Logic, Methodology, and Philosophy of Science*, pp. 240–251. Stanford University Press.

Putnam, H. (1977): 'Realism and Reason'. *Proceedings and Addresses of the American Philosophical Association*, 50(6):483–498.

Putnam, H. (1979): 'Analyticity and Apriority: Beyond Wittgenstein and Quine'. *Midwest Studies in Philosophy*, 4(1):423–441.

Putnam, H. (1980): 'Models and Reality'. *Journal of Symbolic Logic*, 45(3):464–482.

Putnam, H. (1983): 'Philosophers and Human Understanding'. In: *Realism and Reason*, pp. 184–04. Cambridge University Press.

Putnam, H. (1992): 'Replies'. *Philosophical Topics*, 20(1):347–408.

Quine, W. V. (1948): 'On What There Is'. *Review of Metaphysics*, 2(1):21–38.

Quine, W. V. (1949): 'Truth by Convention'. In: Feigl, H. and Sellars, W. (eds), *Reading In Philosophical Analysis*, pp. 250–273. Appleton-Century-Crofts, New York.

Quine, W. V. (1951): 'Two Dogmas of Empiricism'. *Philosophical Review*, 60(1):20–43.

Quine, W. V. (1953): 'On What There Is'. In: *From a Logical Point of View*, pp. 1–19. Harvard University Press.

Quine, W. V. (1960a): 'Carnap and Logical Truth'. *Synthese*, 12(4):350–374.

Quine, W. V. (1960b): *Word and Object*. The MIT Press.

Quine, W. V. (1963): 'Carnap and Logical Truth'. In: Schilpp, P. A. (ed), *The Philosophy of Rudolf Carnap*, pp. 385–406. Open Court, La Salle, Illinois.

Quine, W. V. (1969a): 'Natural Kinds'. In: Kim, J. and Sosa, E. (eds), *Ontological Relativity and Other Essays*, pp. 114–38. Columbia University Press.

Quine, W. V. (1969b): 'Ontological Relaitivity'. In: *Ontological Relativity and Other Essays*, pp. 26–68. Columbia University Press.

Quine, W. V. (1969c): 'Reply to Chomsky'. In: Davidson, D. and Hintikka, J. (eds), *Words and Objections Essays on the Work of W.V. Quine*, pp. 302–311. Reidel, Dordrecht.

Quine, W. V. (1976): 'Whither Physical Objects?' In: Cohen, R. S., Feyerabend, P. K., and Wartofsky, M. (eds), *Essays in Memory of Imre Lakatos*, pp. 497–504. Reidel, Dordrecht.

Quine, W. V. (1981a): 'Things and Their Place in Theories'. In: *Theories and Things*, pp. 1–23. Harvard University Press.

Quine, W. V. (1981b): 'What Price Bivalence?' *Journal of Philosophy*, 78(2):90–95.

Quine, W. V. (1984): 'Book Review: *Mathematics in Philosophy* by Charles Parsons'. *Journal of Philosophy*, 81(12):783–794.

Quine, W. V. (1986a): 'Reply to Charles Parsons'. In: Hahn, L. E. and Schilpp, P. A. (eds), *The Philosophy of W.V. Quine*, pp. 396–404. Open Court, Chicago.

Quine, W. V. (1986b): 'Reply to Hilary Putnam'. In: Hahn, L. E. and Schilpp, P. A. (eds), *The Philosophy of W.V. Quine*, pp. 427–431. Open Court, Chicago.

Quine, W. V. (1990): *Pursuit of Truth*. Harvard University Press.

Quine, W. V. (1992): 'Structure and Nature'. *Journal of Philosophy*, 89(1):5–9.

Quine, W. V. (1994): 'Responses'. *Inquiry*, 37(4):495–505.

Quine, W. V. (1997): *From Stimulus to Science*. Harvard University Press.

Quine, W. V. (2008): *Quine in Dialogue*. Edited by D. Føllesdal and D. B. Quine. Harvard University Press.

Raatikainen, P. (2005): 'On Horwich's Way Out'. *Analysis*, 65(3):175–177.

Raatikainen, P. (2020): 'Gödel's Incompleteness Theorems'. In: Zalta, E. N. (ed), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2020.

Ramsey, F. P. (1925): 'The Foundations of Mathematics'. *Proceedings of the London Mathematical Society*, 25:338–384.

Rayo, A. (2007): 'Logicism Reconsidered'. In: *The Oxford Handbook of Philosophy of Mathematics and Logic*, pp. 203–232. Oxford University Press.

Rayo, A. (2013): *The Construction of Logical Space*. Oxford University Press.

Rayo, A. (2014): 'Reply to Critics'. *Inquiry*, 57(4):498–534.

Richardson, A. (1994): 'Carnap's Principle of Tolerance'. *Proceedings of the Aristotelian Society*, 67:67–82.

Richardson, A. (2012): 'Carnap's Place in Analytic Philosophy and Philosophy of Science'. In: Wagner, P. (ed), *Carnap's Ideal of Explication and Naturalism*. Palgrave-Macmillan.

Richardson, A. W. (1997): *Carnap's Construction of the World: The Aufbau and the Emergence of Logical Empiricism*. Cambridge University Press.

Ricketts, T. (1985): 'Frege, the Tractatus, and the Logocentric Predicament'. *Noûs*, 19(1):3–15.

Ricketts, T. (1986a): 'Objectivity and Objecthood: Frege's Metaphysics of Judgment'. In: Hintikka, J. and Haaparanta, L. (eds), *Frege Synthesized*, pp. 65–95. Reidel, Dordrecht.

Ricketts, T. (1986b): 'Generality, Meaning, and Sense in Frege'. *Pacific Philosophical Quarterly*, 67(3):172–195.

Ricketts, T. (1994): 'Carnap's Principle of Tolerance, Empiricism, and Conventionalism'. In: Clark, P. and Hale, B. (eds), *Reading Putnam*, pp. 176–200. Blackwell, Oxford.

Ricketts, T. (1996): 'Carnap: From Logical Syntax to Semantics'. In: Giere, Ronald, N., and Richardson, A. W. (eds), *Origins of Logical Empiricism*, pp. 231–50. University of Minnesota Press, Minneapolis.

Ricketts, T. (1997): 'Frege's 1906 Foray Into Metalogic'. *Philosophical Topics*, 25(2):169–188.

Ricketts, T. (2003): 'Languages and Calculi'. In: Hardcastle, G. L. and Richardson, A. W. (eds), *Logical Empiricism in North America*, pp. 257–280. University of Minnesota Press, Minneapolis.

Ricketts, T. (2004): 'Frege, Carnap, and Quine: Continuities and Discontinuities'. In: Awodey, S. and Klein, C. (eds), *Carnap Brought Home: The View from Jena*. Open Court, Chicago and LaSalle, Illinois.

Ricketts, T. (2007): 'Tolerance and Logicism: Logical Syntax and the Philosophy of Mathematics'. In: Friedman, M. and Creath, R. (eds), *The Cambridge Companion to Carnap*, pp. 200–225. Cambridge University Press.

Ricketts, T. (2009): 'From Tolerance to Reciprocal Containment'. In: Wagner, P. (ed), *Carnap's Logical Syntax of Language*, pp. 217–235. Palgrave-Macmillan.

Ricketts, T. (2011): 'Roots of Ontological Relativity'. *American Philosophical Quarterly*, 48(3):287–300.

Robinson, A. (1961): *Non-Standard Analysis*. North-Holland Publishing Co., Amsterdam.

Rosser, B. (1937): 'Gödel Theorems for Non-Constructive Logics'. *Journal of Symbolic Logic*, 2(3):129–137.

Russell, B. (1912): *The Problems of Philosophy*. Williams and Norgate, London.

Russell, B. (1918): *Mysticism and Logic and Other Essays*. Longmans, Green & Co., London.

Russell, G. (2008): *Truth in Virtue of Meaning*. Oxford University Press.

Sarkar, S. (2013): 'Carnap and the Compulsions of Interpretation: Reining in the Liberalization of Empiricism'. *European Journal for Philosophy of Science*, 3(3):353–372.

Schmitz, F. (2007): 'Dispositions and Counterfactuals. From Carnap to Goodman's Children and Grandchildren'. In: Kistler, M. and Gnassounou, B. (eds), *Dispositions and Causal Powers*, pp. 43–66. Routledge, London.

Sellars, W. (1963): 'Empiricism and Abstract Entities'. In: Schilpp, P. (ed), *The Philosophy of Rudolf Carnap*, pp. 431–468. Open Court, La Salle, Illinois.

Shapiro, S. (1991): *Foundations Without Foundationalism: A Case for Second-Order Logic*. Oxford University Press.

Shapiro, S. (2000): *Thinking About Mathematics: The Philosophy of Mathematics*. Oxford University Press.

Shapiro, S. (2019): 'Idealization, Mechanism, and Knowability'. In: Horsten, L. and Welch, P. (eds), *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*, pp. 189–207. Oxford University Press.

Sider, T. (2014): 'Outscoping and Discourse Threat'. *Inquiry*, 57(4):413–426.

Simpson, S. G. and Yokoyama, K. (2013): 'Reverse Mathematics and Peano Categoricity'. *Annals of Pure and Applied Logic*, 164(3):284–293.

Skolem, T. (1967): 'Some Remarks on Axiomitized Set Theory'. In: Van Heijenoort, J. (ed), *From Frege to Gödel*, pp. 290–301. Harvard University Press.

Smith, P. (2013): *An Introduction to Gödel's Theorems*. Cambridge University Press.

Stang, N. (2016): 'Kant's Transcendental Idealism'. In: Zalta, E. N. (ed), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2016.

Stanley, J. (1996): 'Truth and Metatheory in Frege'. *Pacific Philosophical Quarterly*, 77(1).

Stein, H. (1992): 'Was Carnap Entirely Wrong, After All?' *Synthese*, 93(1-2):275–295.

Steinberger, F. (2016): 'How Tolerant Can You Be? Carnap on Rationality'. *Philosophy and Phenomenological Research*, 92(3):645–668.

Stroud, B. (1965): 'Wittgenstein and Logical Necessity'. *Philosophical Review*, 74:504–518.

Sullivan, P. M. (2004): 'Frege's Logic'. In: Gabbay, D. M., Woods, J., and Kanamori, A. (eds), *Handbook of the History of Logic*, pp. 659–750. Elsevier, Amsterdam.

Sullivan, P. M. (2005): 'Metaperspectives and Internalism in Frege'. In: Beaney, M. and Reck, E. H. (eds), *Gottlob Frege. Critical Assessments of Leading Philosophers. Volume II: Frege's Philosophy of Logic*, pp. 85–105. Routledge, London.

Tait, W. W. (1986): 'Truth and Proof: The Platonism of Mathematics'. *Synthese*, 69(3):341–370.

Tarski, A., Mostowski, A., and Robinson, R. M. (1953): *Undecidable Theories*. North-Holland Publishing Company, Amsterdam.

Taylor, B. (1991): ''Just More Theory': A Manoeuvre in Putnam's Model-Theoretic Argument for Antirealism'. *Australasian Journal of Philosophy*, 69(2):152–166.

Thomasson, A. L. (2007): *Ordinary Objects*. Oxford University Press.

Thomasson, A. L. (2009): 'Answerable and Unanswerable Questions'. In: Chalmers, D. J., Manley, D., and Wasserman, R. (eds), *Metametaphysics: New Essays on the Foundations of Ontology*, pp. 444–471. Oxford University Press.

Thomasson, A. L. (2015): *Ontology Made Easy*. Oxford University Press.

Turner, J. (2015): 'The Construction of Logical Space and the Structure of Facts'. *Philosophical Studies*, 172(10):2609–2616.

Turquette, A. R. (1950): 'Gödel and the Synthetic a Priori'. *Journal of Philosophy*, 47(5):125–129.

Tymoczko, T. (1989): 'In Defense of Putnam's Brains'. *Philosophical Studies*, 57(3):281–97.

Uebel, T. (1992): 'Neurath vs. Carnap: Naturalism vs. Rational Reconstructionism Before Quine'. *History of Philosophy Quarterly*, 9(4):445–470.

Uebel, T. (2007): *Empiricism at the Crossroads: The Vienna Circle's Protocol-Sentence Debate*. Open Court, La Salle, Illinois.

van Heijenoort, J. (1967): 'Logic as Calculus and Logic as Language'. *Synthese*, 17(1):324–330.

Verhaegh, S. (2018): *Working From Within: The Nature and Development of Quine's Naturalism*. Oxford University Press.

von Neumann, J. (2005): *John Von Neumann: Selected Letters*. American Mathematical Soc.

Wagner, P. (2012): 'Natural Languages, Formal Systems, and Explication'. In: Wagner, P. (ed), *Carnap's Ideal of Explication and Naturalism*. Palgrave-Macmillan.

Waismann, F. (1979): *Wittgenstein and the Vienna Circle: Conversations*. Blackwell, Oxford.

Wang, H. (1996): *A Logical Journey: From Gödel to Philosophy*. The MIT Press.

Warren, J. (2015a): 'Conventionalism, Consistency, and Consistency Sentences'. *Synthese*, 192(5):1351–1371.

Warren, J. (2015b): 'The Possibility of Truth by Convention'. *Philosophical Quarterly*, 65(258):84–93.

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. (forthcoming): 'Infinite Reasoning'. *Philosophy and Phenomenological Research*.

Warren, J. and Waxman, D. (2020): 'A Metasemantic Challenge for Mathematical Determinacy'. *Synthese*, 197(2):477–495.

Williamson, T. (2007): *The Philosophy of Philosophy*. Blackwell.

Williamson, T. (2013): *Modal Logic as Metaphysics*. Oxford University Press.

Wilson, J. M. (2011): 'Much Ado About 'Something''. *Analysis*, 71(1):172–188.

Wittgenstein, L. (1956): *Remarks on the Foundations of Mathematics*. Blackwell.

Wittgenstein, L. (1961): *Tractatus Logico-Philosophicus.* Translated by D.F. Pears and B.F. McGuinness. Routledge.

Wittgenstein, L. (1989): *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge, 1939*. Edited by Cora Diamond. University of Chicago Press.

Woleński, J. (2005): 'Gödel, Tarski and Truth'. *Revue Internationale de Philosophie*, 234:459–490.

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

Wrigley, W. D. (2019): *Axiomatization and Incompleteness in Arithmetic and Set Theory*. Thesis, University of Cambridge.

Yablo, S. (1992): 'Review of Alan Sidelle's *Necessity, Essence, and Individuation: A Defense of Conventionalism*'. *Philosophical Review*, 101(4):878.