A Tale of Two Carnaps

Benjamin Marschall

Draft (March 14, 2022) – Comments welcome: bm515@cam.ac.uk

Abstract

What is Carnap's philosophy of mathematics? And is it any good? Many think that the answers to these question are obvious, but disagree on what they are. According to one still widespread interpretation Carnap is a conventionalist about mathematical truth. His position is consequently regarded as refuted by the influential arguments of Quine and Gödel. Others, however, have argued that Carnap's actual view is more sophisticated and defensible than the conventional wisdom suggests. Allegedly decisive objections fail to meet their target. What explains this extreme divergence? I will show that Carnap's writings allow for two plausible but distinct interpretations: one is more conventionalist, the other more deflationary. On the conventionalist reading Carnap does face some major philosophical challenges, even though they might be surmountable. The deflationary reading does not give rise to this problem, but, on the other hand, struggles to explain a number of argumentative moves Carnap makes. Those who like Carnap's approach to the philosophy of mathematics are thus well advised to decide which reading they want to endorse, and why.

1 Introduction

Carnap's philosophy of mathematics has been controversial. Probably the most contentious issue is the basic question of what his account even *is*. It was (and still is) very common to interpret Carnap as a linguistic conventionalist about mathematical truth (Putnam 1979, Potter 2000: chapter 11, Warren 2020: chapter 13). On this reading it is not surprising that Carnap's position is often regarded as uninteresting and not worth engaging with, at least among systematic philosophers of mathematics.[1]. After all, conventionalism has – allegedly – been refuted by the famous arguments of Quine, and Gödel has some less well-known but powerful arguments too (Quine 1949, Gödel 1995). A very different picture emerges, however, if one looks at the Carnap scholarship that has emerged since

[1] Linnebo's introductory textbook in the philosophy of mathematics from 2017 does not mention Carnap once (Linnebo 2020), for instance, and neither does Colyvan's textbook from 2012 (Colyvan 2012)

the 1980s. Here it has been forcefully argued that Carnap is *not* a linguistic conventionalist, and that objections against his position taken to be decisive actually miss their intended target. On this reading it is surprising that the strength and originality of Carnap's account have been missed for such a long time.[2]

Thanks to the new wave of interest in Carnap we are now in a better position to understand his philosophical project than ever before. Should we therefore dismiss the traditional conventionalist reading of Carnap completely? Should we agree with those who think that *all* arguments against Carnap's position are based on misreadings? This would be a mistake, since the exegetical situation is murkier than has been realised so far. I will argue that, even once the insights of recent scholarship have been taken account, there are still two distinct interpretations of Carnap's philosophy of mathematics available. One of them can be described as a form of *conventionalism*, even though it is a more modest thesis than has usually been ascribed to Carnap. The other reading is more radically *deflationary*. It denies that there is any need for an account of the nature of mathematical truth at all. Both interpretations also face challenges, however: systematic ones in the first case, exegetical ones in the second.

I will proceed by first outlining the *voluntarist* nature of Carnap's overall approach to philosophy in section 2. Unlike many traditional philosophers, Carnap usually does not put forward philosophical *theses* but instead makes *proposals* and *recommendations*. One of the most important insights of recent scholarship is that many objections to Carnap fail precisely because they do not appreciate this aspect of his philosophy. Any credible interpretation of Carnap's philosophy of mathematics must therefore be in line with his voluntarism.

In section 3 I then introduce the core ideas of Carnap's treatment of mathematics. The notion of analyticity plays a key role, and it also makes a conventionalist reading natural: for Carnap mathematics is analytic, and analyticity is truth in virtue of linguistic rules. Some ways to spell out this idea clash with his other commitments. But, in section 4, I will introduce a view called *mild conventionalism* which can plausibly be ascribed to Carnap. Mild conventionalism is interesting and fits what Carnap says, but it also faces some subtle objections. On this interpretation the viability of Carnap's philosophy of mathematics thus remains uncertain.

These problems can be avoided by reading Carnap as a *radical deflationist* about the philosophy of mathematics (section 5). On this reading the notion

[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, and Lavers 2008.

of analyticity plays no genuine explanatory role at all, and Carnap offers no positive account of the nature of mathematical truth. One interesting consequence is that the apparent disagreement between Carnap and Quine about the analyticity of mathematics dissolves itself. A more problematic feature of the deflationary reading, however, is that it seems to clash with some of Carnap's own arguments, in which the analyticity of mathematics is used to *justify* certain assumptions. Friends of this reading thus have some unfinished business as well.

2 Carnap's Voluntarism

2.1 Description and Normativity

Some areas of philosophy, such as ethics and political philosophy, clearly deal with normative questions. But, on the face of it, much of theoretical philosophy is in the business of *describing* various phenomena. Consider the following representative questions from metaphysics, philosophy of language, and the philosophy of mind:

- Are there abstract objects or only concrete entities?
- Do *that*-clauses designate propositions?
- Are mental properties distinct from physical properties?

It seems that any attempt to answer them must involve *factual* claims about the nature of reality, language, and the mind, rather than anything normative such as a recommendation.

The vast majority of Carnap's publications concern topics like logic, metaphysics, the philosophy of mathematics, and the philosophy of science. It is therefore natural to expect him to make descriptive claims about language and reality. And this expectation seems easily confirmed. Consider, for instance, the following passage from *Meaning and Necessity*:

> 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)

On the face of it, Carnap here makes a number of descriptive claims:

- Properties and relations exist.
- Linguistic expressions such as predicates designate these entities.
- Properties and relations are not mental entities.

But appearances are misleading. Many of Carnap's claims that seem descriptive should actually be understood as *normative*. The convenient label *voluntarism* for this general stance was introduced by Jeffrey, who motivates it as follows:

> [Carnap's] persistent, central idea was: "It's high time we took charge of our own mental lives" time to engineer our own conceptual scheme (language, theories) as best we can to serve our own purposes; [...] time to accept the fact that there's nobody out there but us, to choose our purposes and concepts to serve those purposes [...] (Jeffrey 1994: 847)[3]

It is not so clear, however, how this general insight applies to the case at hand. For sure, the descriptive claims just enumerated are naturally *accompanied by* normative claims. Someone who asserts that there are properties and relations presumably also thinks that readers of their treatise *should* come to believe that these entities exist. But how could Carnap's assertions be normative *rather than* descriptive?

In order to understand this we need to take a closer look at Carnap's methodology. It is best illustrates by focusing on the third claim about the ontological status of properties and relations. How does Carnap justify the assertion that they are not mental entities? A passage from "Empiricism, Semantics, and Ontology" (ESO), which contains similar claims about propositions, is helpful:

> 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 [...] occurs [...] shows that the propositions (and their properties, like necessity, etc.) are not subjective. (Carnap 1956a: 210f)

3 Another important early paper stressing the normative nature of Carnap's project is Stein 1992. A comprehensive contemporary discussion can be found in Carus 2007: chapter 11.

Carnap thus adopts a linguistic approach to metaphysical questions. The ontological status of propositions – i.e. their independence of mind and language – is established by relying on the fact that statements about them do not refer to minds and languages.

Which statements are relevant for determining the ontological status of entities though? At this point we need to introduce some machinery from ESO, in particular the notion of a *linguistic framework*. Linguistic frameworks are constructed languages with explicit rules for talking about things, such as physical objects or numbers. In the framework for propositions, Carnap takes one of the relevant rules to be as follows:

> "*p* is a proposition" may be defined by "*p* or not *p*" (Carnap 1956a: 210)

No reference to languages or minds is present here, so Carnap's argument from above goes through.[4]

One might feel unsatisfied with this approach. For it is natural to have a further question about the rules of the proposition framework, or indeed any other framework: namely what *justifies* the acceptance of *these* rules rather than alternative ones. Clearly Carnap thinks that it would be possible to give linguistic rules for propositions that *do* involve references to minds. Why not choose those rather than the rules Carnap adopts?

A traditional philosophical move would be to rely on the nature of propositions, and to argue that they *really are* mind-independent entities. This is of course the opposite of a linguistic approach to metaphysics, since linguistic rules are justified in terms of prior ontological assumptions. Carnap's response to the question of justification is thus radically different: he *denies* that the notion of justification makes sense when applied to the adoption of rules and whole frameworks, since this is not a factual question with objective standard of correctness. Instead, the question whether to adopt a certain framework or not is a *pragmatic* decision (Carnap 1956a: 221).

Carnap effectively thinks of linguistic frameworks as *tools* that serve practical purposes we might have. The task of philosophers is to develop such tools in the hope that they might be useful to others, for instance scientists who are in the business of explaining and predicting empirical phenomena. Like all tools, the frameworks philosophers construct can be criticised in a number of ways.

[4] For a more detailed exposition and defence of Carnap's treatment of ontology in ESO see Ebbs 2017a and Ebbs 2019.

Most prominently, they might turn out to be unhelpful for the purpose at hand. It is likely, for instance, that a framework that does not allow talk of any abstract objects will not suffice to do physics, and if that is one's aim it will be prudent to reject the suggestion to use it.

We wanted to understand how

> Properties and relations are not mental entities

could be a normative rather than a descriptive claim. The essential components are now in place. The first move is to *relativise* such claims to linguistic frameworks and their rules. For Carnap, a more perspicuous representation of the above claim would thus be as follows:

> According to a particular framework F, properties and relations are not mental entities.

In itself this doesn't make the claim any less descriptive. But – so the second move – frameworks can be adopted to achieve certain goals, and in this way normativity enters the picture. For we can make explicit what is being asserted roughly as follows:

> I, Carnap, propose that we accept a framework F whose linguistic rules entail that properties and relations are not mental entities, because such a framework is best suited for a certain purpose P.[5]

Having gone through one example in detail, we can move on to consider why his voluntarism is so important to properly assess Carnap's philosophy.

2.2 Explication and Fit

Recognising the voluntarist nature of Carnap's project is crucial because many objections that would be forceful against descriptive philosophy cease to apply. To illustrate this, suppose a philosopher develops a linguistic framework for doing science which includes ways of speaking that differ considerably from how scientists actually talk. Is this a reason to reject the framework? Not according to Carnap. His aim is to *improve* ordinary ways of speaking, by replacing potentially vague and ambiguous notions by clearer and more explicit ones. He calls this project *explication*. By constructing linguistic frameworks, philosophers give

[5] As Kraut 2016 and Flocke 2020 show, it therefore makes sense to read Carnap as a non-cognitivist about ontology.

explicit rules for the use of certain concepts that can replace similar but more ambiguous concepts of ordinary language. In most cases there will be considerable similarities in how the two concepts are used, but it is decidedly *not* the goal to capture the ordinary usage as closely as possible:

> 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*. (Carnap 1950: 7, my emphasis)

This point cannot be stressed enough. There are a number of philosophical projects which involve, on the one side, the messy and unregimented natural language, and, on the other side, more explicit frameworks, usually involving formal notation. One salient example is formal semantics. Admittedly, there is no unanimous agreement on what exactly semantic theories are trying to account for: the prediction of linguistic utterances, the processing of language in the brain, or something else. But it is clear that formal semantics is first and foremost a *descriptive* enterprise. If a theory makes predictions that clash with the verdicts of competent speakers, then prima that speaks against the theory. In formal semantics the direction of fit thus goes *from* natural language *to* formal system: the latter aim to capture the former.

In books like *Introduction to Semantics* and *Meaning and Necessity*, Carnap provided some of the technical groundwork for the development of formal semantics (Carnap 1942, Carnap 1956b). But his philosophical motivations for these studies were quite different. For sure, Carnap did not deny that a systematic investigation of natural language is possible, but he was not very interested to pursue this project himself (Carnap 1963: 931, 941). The direction of fit Carnap intended was the reverse of that adopted by formal semanticists. He was happy to construct frameworks which do not correspond to any existing speech behaviour, in the hope that people might *change* their ways of speaking. In Carnap's project of explication formal systems thus play a normative rather than as descriptive role.

In his writings, Carnap spends considerably more time describing formal systems than discussing the methodological status of such constructions, and it is thus no wonder that many of his readers missed this important distinction. But it is a crucial one, since, as Carus rightly stresses, the standards of assessment are very different depending on whether formal systems are used in a descriptive or a normative way:

> 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)

Many objections to Carnap's philosophy that seem powerful initially lose their bite once we keep the normative aims of Carnap's project firmly in mind. This can be illustrated using a famous example: Quine's rejection of analyticity (Quine 1951).

How to understand Quine's objection has of course been widely discussed, and I do not want to go into the exegetical debate here.[6] Let us rely on a reading of his complaint that is both straightforward and supported by the text:

(1) For the notion of analyticity to be philosophically respectable, we need to explain the application conditions of the predicate 'analytic' for natural languages in behavouristic terms.

(2) The task described in (1) cannot be achieved.

The justification for (2) is that, in ordinary language, the notion of *analyticity* is not sufficiently clear and established. There may be some paradigm statements whose analyticity is widely agreed on, but the extension of 'analytic' understood in this sense does not cover everything Carnap wants to classify as analytic.

In light of this it is striking that Carnap was inclined to *agree* with Quine's claim (2). In an unpublished response to "Two Dogmas of Empiricism" he writes the following:

> [...] 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)

Like Quine, Carnap therefore thinks that the analytic/synthetic distinction can be drawn *within* a framework by means of explicit rules, but that this distinction does not correspond to any antecedent usage of the terms. Unlike Quine, however, he does not consider this to be a *problem*. What Carnap does is to *recommend* the adoption of a distinction between analytic and synthetic sentences in line with the rules of the proposed framework. For this it does not matter

[6] For an overview see Creath 2007. Some important papers are Creath 1991, O'Grady 1999, George 2000, Gregory 2003, and Hylton 2021.

whether, in natural language, we *already* draw the distinction in the relevant way. Quine's assumption (1), which demands the latter, is incompatible with Carnap's voluntarism and can therefore be rejected.

On this interpretation Quine thus misses his target, since the task he deems impossible is not something Carnap set himself to do anyway. Let me again stress that more sophisticated reading of Quine, on which his arguments retain dialectical force against Carnap, are available. But for our purposes it was sufficient to demonstrate that Carnap's voluntarism makes a big difference to what the criteria of success for his project are. We can now move on to Carnap's philosophy of mathematics, where the notion of analyticity also plays an important role.

3 Carnap's Philosophy of Mathematics

3.1 Mathematics as Analytic

It is difficult to sum up Carnap's philosophy of mathematics in a concise and uncontested way. Even its relationship to established positions such as Platonism, formalism, or logicism is hard to pin down. A commonly applied label is that of a *conventionalist*, however, which amounts to the following view:

> Linguistic Conventionalism
> Mathematical statements are true (or false) in virtue of linguistic rules from which they follow. The adoption of such rules is purely a matter of convention.

This is not too surprising either. As we will see in more detail soon, Carnap sets up his frameworks in such a way that mathematical truths are classified as analytic (Carnap 1937: 116). Analyticity, in turn, is following from the linguistic rules of the framework (Carnap 1956a: 209). Put together this sounds a lot like the conventionalist thesis.

As I have already mentioned, however, conventionalist readings of Carnap have been widely rejected in recent years. To begin with it will therefore be helpful to set out the central tenets of Carnap's philosophy of mathematics in as neutral a way as possible. Once this has been done we can then consider their philosophical significance in more detail.

Carnap's most detailed treatment of mathematics is his *Logical Syntax of Language* (Carnap 1937). The bulk of this book does not contain philosophical discussions, but rather consists in a description of two linguistic frameworks for doing mathematics: language I and language II. Language I is a version of primitive recursive arithmetic (PRA), and language II is a version of the simple theory of types. One result Carnap wants to establish is the following:

> ANALYTICITY
> It is possible to give a definition of 'analytic' for languages I and II such that: each sentence that only contains mathematical vocabulary is either classified as analytic or contradictory, where a sentence is contradictory if and only if its negation is analytic.

This is not a trivial task. Consider language I. One might think that we could just choose an axiomatisation of PRA, and then call those sentences analytic which are theorems of PRA. But this will not do. Carnap wants his definition of 'analytic' to be *complete*: there is supposed to be no purely mathematical sentence which counts as *synthetic*, i.e. neither analytic or contradictory. And since we know from Gödel's incompleteness theorems that there is no complete proof system for PRA, analyticity cannot be equated with theoremhood (Carnap 1937: 41).

Carnap's solution is to use *non-recursive* resources to define 'analytic'. He draws a distinction between two kinds of rules: rules of *derivation*, which are the usual recursive rules used in proof systems, and rules of *consequence*, which include rules that are not recursive. One example of the latter is the following infinitary rule:

ω-RULE

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

Adding these ω-rule to PRA does result in a complete theory. Once rules of consequence are taken into account, Carnap is thus able to define 'analytic' for language I in the desired way (Carnap 1937: 173).

For language II, Carnap's approach is more complicated. He also needs to go beyond the recursive rules of derivation, but since language II is much stronger there is no one infinitary rule that can just be added. Instead, Carnap's approach is *semantic*. While he does not put it that way himself, his definition of 'analytic' for language II is effectively a version of Tarski's definition of truth (Coffa 1987, Koellner ms). We cannot go through all the details here, but one extract will demonstrate that the semantic idea of evaluating a formula relative to a *domain of objects plays* a crucial role. Carnap writes that, 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 1937: 106 (§34c))

This looks like a special case of how the truth-conditions for quantified statements are specified in a modern model-theoretic semantics:

> '$\forall x F(x)$' is true iff, for every element o in the domain of quantification: if o is the object assigned to 'x', then '$F(x)$' is true.

And this analogy is not superficial. While Carnap still describes his definition as syntactic rather than semantic, his infinite sentential class of accented expressions plays the same role as a domain of objects that are being quantified over. Carnap assumes the class of numerals to be isomorphic to the natural numbers, and so analyticity for language II amounts to truth in the standard model of arithmetic:

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

Soon after writing *Logical Syntax* Carnap then officially endorsed Tarski-style semantics and ceased to insist on calling his approach syntactical (Ricketts 1996).

I said that Carnap defines 'analytic' by relying on non-recursive means. Unlike in the case of the infinitary ω-rule used for language I, it might not be immediately obvious in what sense exactly this semantic definition is non-recursive. This question will be of great importance in the next part of the paper, where I present an objection to this aspect of Carnap's philosophy of mathematics. Let us therefore postpone the discussion to section 4.3 and instead focus on the following first: Why did Carnap think that analyticity, characterised in the ways described, is philosophically helpful and important in an account of mathematics? This brings us back to the issue of conventionalism.

3.2 Carnap and Conventionalism

Here are two representative quotes opposing the idea that Carnap is a linguistic 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)

In order to understand this, it is helpful to be more explicit about what conventionalism amounts to. One can distinguish between two ways in which conventions may determine truth. One of them is mundane. Whether a sentence is true or false obviously depends on what the sentence *means*. It is also clear that there is no *natural* connection between sentences and their meanings, since the same content can be expressed using very different words in distinct natural languages. It is thus uncontroversial that conventions play a role in mapping sentences to the propositions they express.

The thesis of linguistic conventionalism is usually understood to involve a different notion of determination than the mundane one. Whereas conventions help to determine the truth *conditions* of all sentences, in some cases, such as mathematics, they also determine the relevant truth *values*. Not only do conventions associate the string "2+2=4" with a certain mathematical proposition, that is, but they also make it the case that this proposition is true (Boghossian 2017).

Ricketts and Ebbs deny that Carnap is a linguistic conventionalist understood in this latter sense. And that is extremly plausible. For the very formulation of the thesis seems to require a metaphysical notion such as truthmaking – something that Carnap would have found too unclear to employ. But does this show that *no* trace of conventionalism remains in Carnap's work? I would hesitate to go this far. Consider the following idea that arguably underlies more metaphysically loaded forms of conventionalism:

> Mild Conventionalism
> In some way or other, the analyticity of mathematics *explains* why mathematics can be accepted without having to deal with traditional philosophical issues concerning ontology, reference, and epistemic access.

Prima facie it is actually quite reasonable to interpret Carnap as a mild conventionalist. For a number of argumentative moves he makes in his writings *do* seem to give the notion of analyticity such an explanatory role. One vivid example concerns ontology. It was (and is) common to think that there must be a deep philosophical difference between logic and mathematics. While logic is completely topic neutral, mathematics is about certain objects, and is thus not independent of the way the world is. Carnap raises this concern in *Logical Syntax*:

> If logic is to be independent of empirical knowledge, then it must assume nothing concerning the *existence of objects*. For this reason Wittgenstein rejected the Axiom of Infinity, which asserts the existence of an infinite number of objects. And, for kindred reasons, Russell himself did not include this axiom amongst the primitive sentences of his logic. (Carnap 1937: 140)

On the face of it this appears to be a problem, since Carnap himself does *not* want to draw a sharp line between logic and mathematics. Here's how he resolves the apparent tension:

> This interpretation has, furthermore, the advantage that a sentence which says that the universe of individuals is infinite is not factual but L-true [= analytic]. Thus the difficulty usually connected with the so-called Axiom of Infinity is here avoided (Carnap 1956b: 86fn9)[7]

The reasoning appears to be this: In the mathematical frameworks Carnap proposes, the axiom of infinity is an analytic statement. Only synthetic sentences describe facts, however, whereas analytic statements are just the consequences of linguistic rules. Despite the initial appearance to the opposite, the axiom of infinity therefore does not commit one to any claims about the nature of reality. And this is exactly the kind of move a mild conventionalist would make.

In addition, even those who want to distance Carnap from conventionalism seem to have no objections to this mild form. In a joint paper with Goldfarb, Ricketts is for instance happy to describe Carnap's position in the following way:

> If the mathematical part of a framework is analytic, then it's analytic; and so invoking mathematical truths at the level of the metalanguage is perfectly acceptable, since they flow from the adoption of the metalanguage. (Goldfarb and Ricketts 1992: 71)[8]

If we take the notion of *flowing* seriously here, the analyticity of mathematics surely does *some* explanatory work – just as mild conventionalism suggests. What we need for more clarity, however, is a way to unpack this metaphor.

7 There is a corresponding but less concise passage in *Logical Syntax* (Carnap 1937: 141). See also Lavers 2016.

8 See also Ebbs 2017b: 26 for an unproblematic notion of "determination" of truth by rules.

In the rest of this paper I proceed as follows. First, I will stick with the working assumption that Carnap is a mild conventionalist. The next section introduces a way to cash out the idea that analytic sentences flow from the adoption of a framework. I then argue that, on this reading, Carnap's philosophy of mathematics faces some serious challenges. Section 5 is devoted to an alternative and more radically deflationary interpretation of Carnap, which rejects mild conventionalism and the explanatory role of analyticity altogether.

4 The Conventionalist Carnap

4.1 Accepting an Explication

Let us begin by reflecting on the status of languages I and II. More specifically, what exactly did Carnap hope to achieve by writing *Logical Syntax*? Keeping his voluntarism firmly in mind, the natural answer is this: Carnap provides explicit proposals for frameworks that we can use in order to do mathematics. They are supposed to be useful for, among others, scientists doing physics. Whether one should prefer to adopt language I, II, or a yet different one depends on a number of factors: how strong a mathematical theory is required, for instance, and how worried one is about the possibility of hidden contradictions.

In sum, Carnap's hope is that people adopt and use and the frameworks of *Logical Syntax*. But what does it mean to *adopt* a framework or an explication? So far I have talked about this activity as if were self-explanatory. But it isn't, and further scrutiny is needed now. This will enable us to flesh out the content of a mild conventionalist reading of Carnap.[9]

Suppose that a Carnapian framework is put forward as an explication of a certain concept. I take it to be obvious that, for someone to adopt the framework, their behaviour must *change* in some way. This can be illustrated by an example Carnap himself gives: temperature. He imagines a community of speakers that have qualitative and comparative temperature concepts (hot, cold, warmer than), but no numerical scale of degrees yet (Carnap 1950: 9f). It is easy to imagine a theorist who proposes the introduction of such a scale, to be used in contexts where precision and quantitative comparisons are desirable. For the community of speakers to adopt this proposal, it is clearly required that they start to *use* the quantitative temperature concepts in real life situations. An explication whose

[9] Unfortunately Carnap himself spends little time discussing the nature of framework adoption. The most salient remarks can be found in Carnap 1939: sections 2-4. I will draw on ideas from Ricketts 2003 and Carus 2007: chapter 11.

adoption does not lead to any differences in speech behaviour has no point.

Roughly, adopting a linguistic framework must therefore be a change in one's linguistic dispositions such that they correspond to the rules of the framework. If one of the adopted framework rules is modus ponens, for instance, we should strive to infer in ways that comply with this inference rule. For a more general definition of framework adoption, the following is a sensible first attempt:

> Disposition
> A person P has adopted a framework F with rules R1, R2, . . . if and only if P is disposed to infer in accordance with rules R1, R2, ... This includes, among other things, a disposition to accept the analytic sentences of F, since they can be derived purely from F's rules.

This definition is still underspecified in many respects. What exactly is it for linguistic behaviour, which consists in uttering sounds or producing marks of ink, to correspond to formal inference rules? What does it take to have certain dispositions? But we need to take some notions for granted to get started, so let us explore how viable this first attempt is.

The second clause of Disposition appears to be too strong. Suppose we have adopted a framework whose inference rules include those of a natural deduction calculus. Then infinitely many logical truths are analytic in this framework. But is anyone really disposed to accept all of these? No, because when faced with a long and complicated logical truth, one is more likely to shrug one's shoulders rather than to assert or deny it.

Nevertheless, there is something sound in the idea that the acceptance of inference rules is not irrelevant to the status of the analytic sentences derivable from them. I think that this connection is best captured in the following way:

> Commitment
> A person P has adopted a framework F with rules R1, R2, ... if and only if P is disposed to infer in accordance with rules R1, R2, ...This includes, among other things, a *commitment* to accept the analytic sentences of F, since they can be derived purely from F's rules.

By a commitment to a sentence S I mean the following: if it were demonstrated to P that S can be derived from F's rules, then P has to accept S.

Further refinements and amendments can be envisaged. At present framework adoption is a rather individualistic endeavour, and it is plausible that someone could be committed to a the rules of a framework in virtue of the fact that the linguistic community they are part of have adopted it. But for the purposes of this paper the formulation Commitment is good enough. Let us now apply it to particular cases.

4.2 Recursive and Non-Recursive Rules

We saw that Carnap's mathematical frameworks – languages I and II – are intended as proposals. We just discussed what the adoption of a framework means in the abstract. Putting these two ideas together, let us ask: what does it take to adopt language I? Among other things, Commitment gives rise to the following requirement:

> In order to adopt language I, one has to be committed to classify each purely mathematical sentence as either analytic or contradictory.

In stating the rules of language I, Carnap made sure to give a definition of 'analytic' that is *complete*: it classifies each purely mathematical sentence as either analytic or contradictory. Since adopting a framework involves accepting the rules that constitute the framework, this brings about the commitment just described.

A commitment to classifying each purely mathematical sentence as analytic or contradictory can be understood in two ways. It might mean that, when faced with a purely mathematical sentence, we are committed to the following claim: this sentence is *either* analytic *or* contradictory. But we are uncommitted about *which* of these options it is. Carnap, however, wants the linguistic rules he gives to do more. They not only settle *that* each mathematical sentence is analytic or contradictory, but also *which* option it is. This emerges especially clearly in Carnap's discussion of language II, where he uses the linguistic rules to show that certain undecidable sentences of the language are analytic rather than contradictory (Carnap 1937: 133). The adoption of one of Carnap's languages should therefore require this more demanding type of commitment. This is also in line with the example of logical truth I used to motivate the idea of commitment earlier. Someone who is shown a proof of a theorem should not merely conclude that it is a theorem *or* a contradiction, but accept it as a theorem.

Thinking about the adoption of language I in this manner is helpful, since it gives substance to the metaphorical claim that mathematical truths "flow from the adoption of the metalanguage". While adopting one of Carnap's frameworks commits one to classifying each mathematical sentence as either analytic or contradictory, there is no such commitment for *synthetic* sentences that contain non-logical or non-mathematical vocabulary. For sure, Carnap's frameworks use classical logic, and so those who adopt them will, for each synthetic sentence, be committed to the following disjunctive claim: this sentence is either true or false. But for the mathematical part of the language the commitment is much more informative. It also settles the truth *values* of analytic sentences, whereas empirical investigations are required for this in the synthetic case. There is thus a clear sense in which mathematical truth is fundamentally different from empirical truth.[10]

The current proposal is a way to spell out the mild conventionalist reading of Carnap that does not rely on metaphysical notions he would reject at the outset. Most importantly, it is compatible with Carnap's voluntarism. The core idea is not, after all, that we are in some sense *already* committed to classifying mathematics as analytic. Rather, Carnap is proposing a kind of language reform: if we like his language I or II, we should adopt them by forming dispositions to use their rules, which then give rise to the relevant commitments. Carnap's formal systems thus have a normative role, as desired.

However, there are also some less good news. As I will argue next, there is reason to doubt whether it is actually *possible* to adopt language I. And if this suspicion is correct, then Carnap is in trouble. Showing that a certain proposal is impossible to use in practice surely is among the most serious criticisms one can raise against it.

Why think that there is any problem with adopting language I? Remember that, in order to overcome Gödel's limitative results, Carnap had to rely on the non-recursive, infinitary ω-rule. This was an unusual and contentious move already at the time of writing *Logical Syntax*, but Carnap was undeterred:

> 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 1937: 173)[11]

The last claim is doubtful. For while we can make good sense of what it means to adopt and use *recursive* inference rules, the situation is different for a rule that,

[10] One can understand the mild conventionalist Carnap as an *inferentialist* about mathematical truth. Compare the position defended in Warren 2020.

[11] As Buldt 2004 shows, Carnap initially shared the widespread skepticism about the ω-rule when he first learned about it through Hilbert and Gödel.

like the ω-rule, has infinitely many premises. As finite beings in a finite world, we will never actually encounter infinitely many premises to draw a conclusion from. So in what sense could we decide to follow this rule?[12]

To make this worry more concrete, consider some mathematical sentence ϕ of language I that can only be derived using the ω-rule. The adoption of language I should commit us to ϕ being analytic. But what does this commitment really amount to? In this case we cannot imagine that someone shows us a proof of ϕ using rules we accept, since this infinite proof could not be written down. It is thus natural to conclude that it is only possible to adopt the *recursive* rules of a framework, not the non-recursive ones.

The practical application of infinitary rules has received little attention in the secondary literature on Carnap. Ricketts explicitly addresses this point, and seems to endorse the conclusion just reached:

> It is implausible to hold that a classical mathematician [...] is in any sense disposed to affirm or deny each mathematical sentence of this language. [...] Speech habits thus do not fix the [non-recursive] L-rules for a calculus instantiated by a language. (Ricketts 2003: 262)

Ricketts does not intend this observation to be an *objection* to Carnap's approach. But, on the mild conventionalist reading, it is one. For if speech habits do not suffice to commit us to the non-recursive rules of a framework, then there is no commitment to the analyticity of mathematics. We thus loose what we seemed to have gained: a way to elucidate the idea that mathematical truth flows from the adoption of rules.

One response on behalf of those who like the mild conventionalist reading of Carnap involves going on the offensive. Contrary to what I have suggested, they might argue that we *can* follow the infinitary ω-rule. This is no easy task, but some recent work suggests a promising way to defend a Carnapian form of conventionalism against the conventional wisdom (Warren 2020, Warren forthcoming).

In what follows I will address a different response, however, according to which my focus on language I and the ω-rule has been misdirected energy. This, supposedly, is because for language II, Carnap uses a semantic strategy to define 'analytic'. And this approach, on the face of it, does not involve any infinitary rules that human beings are unable to follow. The strategy is also more general

12 For this reason philosophers of mathematics tend to side against Carnap and deny that following the ω-rule is possible (Field 1994, Raatikainen 2005, Smith 2013: 332, Button and Walsh 2018: chapter 7).

in scope. So, one may wonder, does this not render the current objection moot? I think not, but the matter requires a closer look which we will turn to now.

4.3 Tarski to the Rescue?

As has been said before, Carnap's definition of 'analytic' for language II is effectively a notational variant of a definition of truth in the style of Tarski. On the face of it, there appears to be nothing infinitary about a Tarskian truth definition. Its key feature is a finite number of clauses which specify the truth conditions for different types of sentences. Here's one simplified way to spell this out, covering only monadic predicates:

- Atomic sentences: 'Fx' is true iff the object assigned to 'x' is in the extension assigned to 'F'
- Negated sentences: '$\neg\phi$' is true iff 'ϕ' is false
- Conjunctions: '$\phi \wedge \psi$' is true iff 'ϕ' is true and 'ψ' is true.
- Universal generalisations: '$\forall x F(x)$' is true iff, for every o in the domain of quantification: if o is the object assigned to 'x', then '$F(x)$' is true.

Using recursion, this definition provides truth conditions for arbitrarily complex sentences.

Do I want to claim that, like with the ω-rule, there is some deep puzzle about how we can adopt the Tarski rules in practice? No. I agree that this is perfectly possible. What I will argue for instead is that, when Carnap is read as a mild conventionalist, accepting his definition of analyticity involves *more* than just accepting the recursive clauses.

In and of themselves, the Tarski clauses just state truth *conditions*. Take the claim that all things are red, formalised as '$\forall x Rx$'. This is true if all objects in the domain of quantification are in the extension of the predicate 'R'. But whether this is the case of course depends on what the domain of quantification is, and what extension 'R' has. For our purposes, it will be crucial to scrutinise the following question: What, on Carnap's view, is the theoretical role of the domain of quantification?

In *Logical Syntax*, it is clear that Carnap takes the quantifiers of language II to range over a domain that includes accented expressions, i.e. '0' followed by a finite number of $'$s: $0'$, $0''$, and so on. The domain is thus isomorphic to the standard natural numbers. And this assumption is very important. Carnap for instance relies on it when arguing that the consistency sentence of language II is analytic even though underivable. As Beth pointed out, this argument does not go through if one takes the domain of accented expressions to include additional, non-standard numerals (Beth 1963, Marschall 2021).

Where does the particular domain of quantification Carnap relies on come from? As Frost-Arnold rightly stresses, Carnap considers it to be part of the framework rules:

> 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)

However, what kind of activity is specifying a universe of discourse? In some contexts it is easy to understand what is meant by this, such as when we are dealing with quantifier domain restriction. Suppose someone says that "there is no more beer". In most contexts this will be true if there is no more beer *in the fridge*, say, rather than *anywhere in the universe*. This can be explained by holding that the speaker restricts the domain their quantifiers range over to objects in the immediate vicinity. This process can be called domain specification, and it also makes sense to describe it as a linguistic rule. But, as I will argue next, the conventionalist Carnap actually needs something stronger and more problematic.

Consider first what, according to the mild conventionalist reading, it takes to adopt Carnap's language II:

> In order to adopt language II, one has to be committed to classify each purely mathematical sentence as either analytic or contradictory.

If specifying the domain for mathematical discourse is part of the framework rules, then it should give rise to the commitment described. But does it? It is not straightforward how. For comparison, consider a framework for doing physics. One of its rules, presumably, would specify that the intended domain of quantification are all the spatiotemporal entities. But, clearly, adopting this framework does *not* commit us to particular truth *values* for all physical sentences, but only to them having particular truth *conditions*. In order to determine the truth values, further empirical evidence is needed.

For empirical discourse this is the desired result. But on our current reading of Carnap, mathematics is supposed to be different: the truth values – analytic or contradictory – should be consequences of the framework adoption itself. And in order to account for this, the specification of the domain *for the mathematical part of the language* must be a different process than for the empirical part.

The special status of the mathematical domain can be summed up using the following slogan: whereas in the empirical case the domain is specified *by ostension*, in the mathematical case it is specified *by description* (Button and Walsh 2018: 151). Here's what I mean by this. In the empirical case, specifying the domain does not entail that, through the act of specification, we know *which* objects the domain contains. We do know that, for instance, we want to quantify over beers in the house, but *how many* such beers there are is determined by the way the world is, not the specification on its own. In the mathematical case, on the other hand, the specification must be more informative. It is a description of the domain we want to quantify over, such that, for instance, how many numbers there are is not a further fact settled by the way the world is. In other words, the *content* of the domain of quantification must follow from the specification itself.

If specifying a mathematical domain by description were possible, then mild conventionalism would be defensible. Since the specification of the mathematical domain itself provides all relevant information about the mathematical objects, it is clear how the specification could settle the truth values of mathematical claims. But this project faces a problem that bring us back to non-recursive rules. Consider how Carnap actually describes the intended domain for mathematics in *Logical Syntax*: by enumerating accented expressions 0, $0'$, $0''$, and so on. This is clearly not sufficient. In the paper I already alluded to, Beth shows that this description of the domain can easily be understood in an unintended way that includes more numerals than Carnap wants, which gives rise to undesired results (Beth 1963: 478). In order to properly characterise the intended domain, Carnap thus needs a way to single out the standard numerals to the exclusion of non-standard ones.

At this point Gödel's limitative results strike again. For our task is equivalent of finding a theory that only has models isomorphic to the natural numbers, and one consequence of Gödel's incompleteness theorems is that no recursive theory can achieve this. We would thus have to rely on non-recursive resources after all, and are thus back at the challenge raised at the end of the previous section.[13]

We started this section with the suggestion that Carnap's semantic definition of analyticity avoids a commitment to non-recursive methods. I have argued against this by pointing out that, on the conventionalist reading, adopting the

[13] An excellent summary of the technical results relied on here, as well as their philosophical significance, is given in Button and Walsh 2018: chapter 7.

recursive clauses is not enough. Carnap also needs to give the domain of quantification of the mathematical part of the language a special status. And this second task requires non-recursive methods after all, since he needs to single out the standard numerals by description.

Once again, I don't want to claim that it is *impossible* to defend the use of non-recursive methods by the conventionalist Carnap. But whether the position is stable is a substantial question, and further work is needed to be sure. On the current reading, Carnap's voluntarism thus does not suffice to defend his philosophy of mathematics.

5 The Deflationary Carnap

5.1 Analyticity without Conventionalism

The preceding discussion was based on ascribing the following thesis to Carnap:

> Mild Conventionalism
> In some way or other, the analyticity of mathematics *explains* why mathematics can be accepted without having to deal with traditional philosophical issues concerning ontology, reference, and epistemic access.

Some readers will likely have become impatient already, since they think that even mild conventionalism is *too much* conventionalism. They might say that the worries about infinitary rules I raised do not demonstrate any problem Carnap has, but show that we should thoroughly separate his position from all conventionalist ideas. This alternative interpretation must be taken seriously, so let us explore its prospects.

In order to have a label at hand to contrast mild conventionalism with, I will call this reading *radical deflationism*. Its key assumption can be summed up as follows:

> Radical Deflationism
> By constructing linguistic frameworks, Carnap does not aim to answer traditional philosophical questions about mathematics in an *alternative* way. The analyticity of mathematics, for instance, is not an account of mathematical truth in competition with, say, a Platonist view. Instead, Carnap thinks that we can and use mathematical

> frameworks without addressing these philosophical issues at all, and urges us to follow his lead.

The deflationary Carnap is much more opposed to traditional philosophy of mathematics than the conventionalist one. To illustrate this, consider a mathematical Platonist like Gödel. Gödel and the conventionalist Carnap can agree that the nature of mathematical truth is puzzling and requires philosophical elucidation. Gödel's approach involves components Carnap would reject, such as a faculty of mathematical intuition. But the conventionalist Carnap at least addresses the same *question*. Instead of making dubious metaphysical assumptions, he just tries to account for the nature of mathematics by more mundane means, such as rules that we follow.

The deflationary Carnap, on the other hand, doesn't really engage with Gödel head on. His approach is more quietist. Instead of *solving* problems in the philosophy of mathematics in an un-metaphysical manner, he wants to *dissolve* them, by undermining the presupposition that there are any deep puzzles to be addressed in the first place.

The deflationary Carnap can conveniently dodge the critical questions I raised in the previous section. The working assumption there was that Carnap wants to give the idea that "mathematical truth flows from the adoption of the metalanguage" some determinate content, since the notion of analyticity plays an explanatory role. But the deflationary Carnap does not share this goal of his conventionalist counterpart, and therefore he does not need trouble himself over issues like the practical application of infinitary rules.

The deflationary reading gives rise to other pressing questions though. If the notion of analyticity plays no explanatory role in Carnap's philosophy of mathematics, why did he care about it at all? Doesn't the fact that he spent so much time and effort on defining analyticity in itself show that he took the notion to be philosophically important? At the end of the day I think that deflationary readers will indeed struggle to make sense of all the things Carnap says about analyticity. But there is no easy refutation of deflationism either. For there is a natural and straightforward explanation for why even the deflationary Carnap has a place for the notion of analyticity, which I will outline now.

The task of philosophers, so Carnap, is to give explications by proposing linguistic frameworks for adoption. These frameworks come with explicit rules. Hence a Carnapian philosopher will fill their papers and books with formal systems and discussions of their rules, as Carnap himself did. But those rules and what follows from them *just are* the analytic statements of the respective frameworks. In this context, what would it mean to proceed *without* the notion of analyticity? It seems that the only way to do so would be to stop constructing formal systems altogether, and to give up the task of Carnapian explication. Understood in this sense, the notion of analyticity is a *practical necessity*: one cannot describe a formal system without giving any rules, and so one cannot do without analyticity.

There are exceptions to this rule, for one can imagine a Carnapian framework in which no sentence is analytic. The framework rules, in such a case, would only contain formation rules about which sentences are syntactically well-formed, but nothing further. But it is plausible to suppose that such extreme cases have little practical relevance, and thus do not distract from the general lesson above.

Is this really all Carnap meant by calling mathematics analytic? In order to fully appreciate how deflated the notion of analyticity has become on the current interpretation, it will once again be helpful to compare it to Quine's position.

5.2 Quine versus Carnap

I already said that the Carnap-Quine debate over analyticity is notoriously complex, and this is so even if we restrict our attention to the analyticity *of mathematics*. Helpfully for our purposes, in a retrospective article Quine sums up his view on the nature of their disagreement 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 [...] the notion of analyticity ceases to be vital to epistemology. (Quine 2008: 26f)

Quine assumes that Carnap uses the notion of analyticity to address a certain philosophical question: how to make mathematics, with its apparent commitment to abstract objects and an a priori epistemology, compatible with empiricism. Quine agrees that this is a good question, but offers a different cure: holism. Quine's Carnap is thus clearly the *conventionalist* Carnap, not the deflationary one. Quine thinks that his holist approach provides a more satisfactory answer to some genuine philosophical questions than Carnapian analyticity. This kind of challenge has no force against the deflationary Carnap.

That Quine interprets Carnap as a conventionalist is further supported by looking at his paper "Carnap and Logical Truth". To be fair, in the original version Quine starts the article by voicing some doubts about whether his objections actually apply to anything Carnap himself holds:

> 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)

This striking preamble suggests that Quine sensed that Carnap's approach to philosophy was more deflationary and quietist than his own. In the actual text, however, Quine considers Carnap's approach to logic and mathematics to be a philosophical *competitor*. One of Quine's most interesting objections to Carnap's use of analyticity in mathematics can be found in section VII. In reference to Carnap's definition of 'analytic' for language II, Quine writes this:

> 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 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)

As I understand this passage, it contains an interesting objection to the conventionalist reading of Carnap's project. Quine makes the point that one can give a Tarski-style semantics for *any* part of the object language, whether mathematical or empirical, provided that the metalanguage is powerful enough. Carnap wants to distinguish between truth and and falsity for *synthetic*, empirical sentences, on the one hand, and analytic truths and contradictory falsehoods for mathematical statements, on the other hand. But the definition of analyticity takes the form of a truth theory that can also be given for other parts of the language. It is therefore difficult to see why we should not rest content with the notions of truth and falsity *simpliciter*, rather than making the distinction between the analytic and the synthetic.

Quine's preferred solution is to apply 'true' and 'false' to *all* sentences, whether mathematical or empirical. This would be a problem for the conventionalist

Carnap, for whom the analyticity of mathematics is supposed to *explain* something about the nature of mathematical truth. He needs to find a way to resist Quine's conclusion. But for the deflationary Carnap, none of this is any reason to worry. For him the analytic/synthetic distinction is not a distinction between two different *kinds of truth*: factual versus mathematical truth. The analyticity of mathematics is not supposed to explain any puzzling features of mathematical discourse, but is a matter of *presentation*. Since Carnap's aim is construct frameworks for doing mathematics, he obviously needs to state rules for using mathematical vocabulary – and stating rules *just is* giving analytic statements.

One salient question, then, is whether Quine would have objected to the notion of analyticity used by the deflationary Carnap. The answer is *no*. He would be happy to admit that, understood in this sense, analyticity is a coherent notion. To be fair, in some of his early critiques Quine sounds rather determined to argue that the notion of analyticity is wholly useless or even meaningless, but in his later writings a more subtle picture emerges. His considered position is that the notion of analyticity cannot do the *epistemological* work he took Carnap to want it to do, but may have other uses that are perfectly sensible. One such use is the *introduction of new terms* through definitions. About the technical term 'momentum', for instance, Quine writes the following:

> When in relativity theory momentum is found to be not quite proportional to velocity, despite its original definition as mass times velocity, there is no flurry over redefinition or contradiction in terms, and I don't think there should be. The definition served its purpose in introducing a word for subsequent use, and the word was thereafter ours to use in the evolving theory, with no lingering commitments. Definition is episodic. [...] In short, I recognize the notion of analyticity in its obvious and useful but epistemologically insignificant applications. (Quine 1991: 271)

According to this story, the term 'momentum' was introduced through a stipulative definition in terms of the antecedently understood notions of mass and velocity. It is therefore fine to say that, at the time of introduction, the following sentence was analytic:

(A) Momentum = Mass x Velocity

However, the analyticity of (A) only means that it was used to introduce this term. (A) was of instrumental value, but it has no special epistemic significance.

When later scientific development showed that (A) needs to be revised, this is not a philosophical puzzle, since on the current conception analytic sentences do not state eternal truths. They rather introduce notions which can then evolve in various directions.

The definitional understanding of analyticity that Quine accepts, however, is just what the deflationary Carnap means by analyticity as well. The point of giving linguistic rules is that people start to use the framework thereby defined. This does not mean that thereafter the rules can never be revised. Carnapian explication is supposed to be a continuous process of improvement, after all, and one should not assume that the framework that bests suits our current purposes will continue to do so indefinitely. So whereas there is a real conflict between Quine and the conventionalist Carnap, the positions of Quine and the deflationary Carnap are perfectly compatible.

5.3 Is This All?

The deflationary reading of Carnap does very well. It avoids the objections levelled against the conventionalist reading. It also fits the spirit of Carnap's voluntarism. Should we thus forget about the conventionalist Carnap completely, and set him aside as being wholly based on misreadings? That would be premature. For there are passages in which Carnap clearly sounds like he *does* take the notion of analyticity to play an explanatory philosophical role, one that is incompatible with the deflationary reading. I already quoted an example of this in section 3.2: Carnap's treatment of the axiom of infinity. I will now introduce a further and especially striking remark from an unpublished letter. Based on it I will argue that, just like the conventionalists, the deflationary readers of Carnap have outstanding work to do in order to defend their preferred reading.

To start with, think back to the way Carnap defined 'analytic' for languages I and II. Both approaches are technically sophisticated, and Carnap devotes a lot of space in *Logical Syntax* to discussing them. Here's a question a deflationary reader of Carnap should say something about: why did Carnap think that all this ingenuity is required?

Suppose that, instead of using infinitary inference rules and Tarskian machinery, in his specification of languages I and II Carnap had merely said the following:

> Any sentence that consists solely of logical and mathematical vocabulary is to be classified as either analytic or contradictory. In many

> cases we can, in principle, find out which of the two options it is, by deriving the sentence or its negation from the axioms. For undecidable cases this is not possible. In some cases we will remain agnostic, in others there is indirect evidence. Consistency sentences are for instance supported by the fact that in well-studied systems no contradictions have been found.

Would this have been any worse than what Carnap actually said? The conventionalist readers of Carnap will say so. They want the linguistic rules to *determine* whether sentences are analytic or contradictory in some way, so agnosticism is not an option. But does the deflationary Carnap have analogous reasons to demur? It is hard to see why. Since he does not want to retain the conventionalist idea that mathematical truth flows from linguistic rules, it seems that the rule above would be a suitable explication of the analyticity of mathematics as well.

Why did Carnap nevertheless proceed in the way he did? One obvious answer is that he was interested in logic and mathematics as such, and liked to proof and define things. Even though the definitions he gives don't serve the philosophical purpose associated with conventionalism, they are nevertheless interesting and original from a technical point of view. It is thus no wonder that he involved himself with the ω-rule and Tarskian semantics.

I agree that Carnap's technical interests are *part* of the explanation, but am less confident that they are the full story. As I will describe in some detail now, there is textual evidence that Carnap took these technical results to be philosophically significant in a way that fits the conventionalist reading much better than the deflationary one.

In 1949, Copi published a paper called "Modern Logic and the Synthetic A Priori" in the *Journal of Philosophy* (Copi 1949). Copi argues that undecidable mathematical sentences like the Gödel sentence should be classified as *synthetic* but nevertheless a priori, just as Kant suggested for all of mathematics. This spawned a short debate. Turquette replied with the counter-suggestion that undecidable sentences have an *indeterminate* truth value (Turquette 1950), and Copi then doubled back on his original proposal (Copi 1950). Carnap also took part, though not publicly. He wrote a letter to Copi criticising his proposal, which led to some further exchanges.

In his letter, unsurprisingly, Carnap recommends his own approach according to which all of mathematics, including undecidable sentences, is analytic (or contradictory). And the way he phrases his dissent is very interesting. Consider, for instance, the following 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, my emphasis)[14]

The last sentence strongly suggests that, for Carnap, infinitary rules *do* play a philosophically important role: they apparently enable us to *know* that certain undecidable sentences are analytic rather than contradictory. This impression is further supported by what Carnap writes about the Gödel sentence:

> Gödel's sentence is a universal sentence of such a kind that every substitution instance of it (with the constant of a natural number substituted for the variable) is provable and therefore generally recognized as analytic. *The universal sentence does not say anything more than the totality of the instances*. (Carnap, Letter to Copi from August 16 1950, RCP 027-03-07, my emphasis)

In making the step from the analyticity of the infinitely many instances to the analyticity of the universal claim, however, Carnap implicitly needs to rely on something like the ω-rule, as discussed in section 4.2. And if that is so, then, contrary to what was promised by deflationary readers, we *do* need to worry about things like how to use infinitary rules after all.

It must be admitted that the quoted remarks are open to interpretation. For instance, does Carnap merely want to say that *for the Gödel sentence specifically* we are able to determine its analyticity based on infinitary rules? Or is the claim rather something stronger, namely that, *in general*, infinitary rules can guide us to determine whether undecidable sentences are analytic or contradictory? This makes a big difference. Whereas the latter reading pushes us towards conventionalism, the former can probably be integrated into a deflationary reading.

Further investigations are thus needed, but I will stop here. While both the conventionalist and the deflationary readings have their strengths and weaknesses, we can conclude that they both struggle with infinitary rules. The conventionalist Carnap needs to make sense of infinitary rule following, which is

[14] The Rudolf Carnap Papers are based at the Archives of Scientific Philosophy of the University of Pittsburgh.

difficult. The deflationary Carnap avoids this problem, but doesn't have a positive role for infinitary rules at all, which seems contrary to Carnap's own statements. There is work to be done on both sides.

6 Conclusion

I have argued that, even if we keep Carnap's voluntarist conception of philosophy firmly in mind, there are nevertheless two exegetically plausible but distinct interpretations of his philosophy of mathematics: a more conventionalist and a more deflationary one. One question one can ask is which of these, if any, Carnap's *actual* position was. But the spirit of Carnap's philosophy is better captured by a different line of inquiry: which of these readings results in the more *interesting* and *fruitful* position? I have said little about this, but a case can be made for keeping *both* Carnaps around.

The conventionalist Carnap has recently inspired some interesting systematic work on conventionalism in the philosophy of mathematics (Warren 2020). Even though the association of Carnap with conventionalism has done a lot of harm to his reception in the past, it therefore seems unwise to abandon this connection completely. One intriguing feature of the deflationary Carnap, on the other hand, is his apparent compatibility with Quine's position. In my view Quine's own philosophy of mathematics is not very well understood, and the comparison with Carnap should be a helpful entry point. Peaceful coexistence between the two Carnaps might thus be the best approach.

References

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.

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. A. (2017): 'Analyticity'. In: *A Companion to the Philosophy of Language*, pp. 578–618. John Wiley & Sons, Ltd.

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.

Button, T. and Walsh, S. (2018): *Philosophy and Model Theory*. Oxford University Press.

Carnap, R. (1937): *The Logical Syntax of Language*. K. Paul, Trench, Trubner & Co., London.

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. (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. (1963): 'Replies and Systematic Expositions'. In: Schilpp, P. A. (ed), *The Philosophy of Rudolf Carnap*, pp. 859–1013. Open Court, La Salle, Illinois.

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.

Carus, A. W. (2007): *Carnap and Twentieth-Century Thought: Explication as Enlightenment*. Cambridge University Press.

Coffa, A. (1987): 'Carnap, Tarski and the Search for Truth'. *Noûs*, 21(4):547–572.

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. (2007): 'Quine's Challenge to Carnap'. In: Friedman, M. and Creath, R. (eds), *The Cambridge Companion to Carnap*, pp. 316–335. Cambridge University Press.

Ebbs, G. (1997): *Rule-Following and Realism*. Harvard University Press.

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.

Field, H. (1994): 'Are Our Logical and Mathematical Concepts Highly Indeterminate?' *Midwest Studies in Philosophy*, 19(1):391–429.

Flocke, V. (2020): 'Carnap's Noncognitivism About Ontology'. *Noûs*, 54(3):527–548.

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.

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. (1995): 'Is Mathematics Syntax of Language?' In: *Collected Works. Volume III. Unpublished Essays and Lectures.*, pp. 334–355. Oxford University Press: Oxford.

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.

Gregory, P. A. (2003): ''Two Dogmas'–All Bark and No Bite? Carnap and Quine on Analyticity'. *Philosophy and Phenomenological Research*, 67(3):633–648.

Hylton, P. (2021): 'Carnap and Quine on Analyticity: The Nature of the Disagreement'. *Noûs*, 55:445–462.

Jeffrey, R. C. (1994): 'Carnap's Voluntarism'. In: Prawitz, D., Skyrms, B., and Westerståhl, D. (eds), *Logic, Methodology and Philosophy of Science*, pp. 847–866. Elsevier: Amsterdam.

Koellner, P. (ms): 'Carnap on the Foundations of Logic and Mathematics'. Unpublished manuscript. Online: http://logic.harvard.edu/koellner/CFLM.pdf.

Kraut, R. (2016): 'Three Carnaps on Ontology'. In: Blatti, S. and Lapointe, S. (eds), *Ontology After Carnap*, pp. 31–58. Oxford University Press.

Lavers, G. (2008): 'Carnap, Formalism, and Informal Rigour'. *Philosophia Mathematica*, 16(1):4–24.

Lavers, G. (2016): 'Carnap's Surprising Views on the Axiom of Infinity'. *Metascience*, 25(1):37–41.

Linnebo, Ø. (2020): *Philosophy of Mathematics*. Princeton University Press.

Marschall, B. (2021): 'Carnap and Beth on the Limits of Tolerance'. *Canadian Journal of Philosophy*, 51(4):282–300.

O'Grady, P. (1999): 'Carnap and Two Dogmas of Empiricism'. *Philosophical and Phenomenological Research*, 59(4):1015–1027.

Potter, M. (2000): *Reason's Nearest Kin: Philosophies of Arithmetic From Kant to Carnap*. Oxford University Press.

Putnam, H. (1979): 'Analyticity and Apriority: Beyond Wittgenstein and Quine'. *Midwest Studies in Philosophy*, 4(1):423–441.

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. (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. (2008): *Quine in Dialogue*. Edited by D. Føllesdal and D. B. Quine. Harvard University Press.

Quine, W. v. O. (1991): 'Two Dogmas in Retrospect'. *Canadian Journal of Philosophy*, 21(3):265–274.

Raatikainen, P. (2005): 'On Horwich's Way Out'. *Analysis*, 65(3):175–177.

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. (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. (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.

Smith, P. (2013): *An Introduction to Gödel's Theorems*. Cambridge University Press.

Stein, H. (1992): 'Was Carnap Entirely Wrong, After All?' *Synthese*, 93(1-2):275–295.

Turquette, A. R. (1950): 'Gödel and the Synthetic a Priori'. *Journal of Philosophy*, 47(5):125–129.

Warren, J. (2020): *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. Oxford University Press.

Warren, J. (forthcoming): 'Infinite Reasoning'. *Philosophy and Phenomenological Research*.