Carnap, the Conventionalist

Benjamin Marschall

Draft (December 19, 2022) – Comments welcome: `bm515@cam.ac.uk`

Abstract

Conventional wisdom has it that Carnap is a linguistic conventionalist about mathematical truth. This interpretation is natural because Carnap holds that mathematical statements are analytic, where analyticity amounts to following from linguistic rules that can be adopted based on pragmatic considerations. In recent scholarship, however, the traditional conventionalist interpretation has been unpopular. It has been suggested that Carnap's unique metaphilosophy, which is based on the principle of tolerance and the method of explication, excludes a conventionalist reading of his philosophy of mathematics. Against this, I defend the conventional wisdom by arguing that Carnap's position should be understood as what I call *normative commitment conventionalism*. Unlike some versions of conventionalism, it does without the metaphysical notion of truth-in-virtue-of. Unlike some versions of conventionalism, it is not a descriptive thesis but a proposal. It retains the sprit of conventionalism, however, according to which linguistic rules suffice to explain mathematical truth. Admittedly, my conventionalist reading comes with a downside: it requires a form of infinitary rule-following that many have found questionable. Nevertheless, so I argue, it makes better sense of Carnap's views on the analyticity of mathematics than competing interpretations.

1 Introduction

What is Carnap's philosophy of mathematics? Radically different answers have been proposed. It was (and still is) very common to interpret Carnap as a *linguistic conventionalist* about mathematical truth (Putnam 1979, Creath 1992 Potter 2000: chapter 11, Ben-Menahem 2006: chapter 5, Warren 2020: chapter 13). Given this reading, it is not surprising that Carnap's position is widely regarded as not worth engaging with, at least among systematic philosophers of mathematics.[1] After all, conventionalism is often said to have been refuted by Quine's famous

[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). Earlier and more brutally, John von Neumann called Carnap's position "naive and feeble" (von Neumann 2005: 203)

"Truth by Convention" (Quine 1936).[2] However, a new wave of Carnap scholarship that has emerged since the 1980s paints a very different picture. Here it has been forcefully argued that Carnap is *not* a linguistic conventionalist in any straightforward sense, and that objections against his position previously taken to be decisive actually miss their intended target.[3]

Thanks to this renewed interest in Carnap we are now in a better position to understand his philosophical project than ever before. But fundamental questions about his philosophy of mathematics remain unresolved. Should we conclude, as some of the recent scholarship suggests, that there is *no* affinity between Carnap's position and linguistic conventionalism at all? I will argue that this would be a mistake. The secondary literature has neglected to distinguish between different forms of linguistic conventionalism. One form relies on metaphysical notions like *truthmaking*, and it is rightly rejected by the new wave Carnap scholars. A different form, however, only requires the notion of *commitment*. I will argue that it is not only unobjectionable, but even required to make sense of Carnap's well-known idea of linguistic frameworks. His philosophy of mathematics can thus be described as *commitment conventionalism*. I furthermore distinguish between a *descriptive* and a *normative* flavour of this position, and argue that Carnap accepts the normative version.

I proceed as follows. In the next section I begin by sketching Carnap's views on mathematics in as neutral a way as possible. His reliance on the notion of analyticity fuelled the idea that Carnap is a conventionalist about mathematical truth, and I will distinguish between a metaphysical and a commitment-based version to spell out this thought. After dismissing the metaphysical version, section 3 contains the first step of my positive argument. I show how Carnap's metaphilosophy, which is based on the principle of tolerance and the method of explication, naturally leads to what I call normative commitment conventionalism. My argument, however, depends on an assumption that has been contentious: that there are facts about whether someone is a speaker of a particular language. In a second step, section 4 thus defends this assumption against an alternative interpretation put forward by Ricketts, according to which Carnap denies that there are such facts. Section 5 is then devoted to a critical assessment of Carnap's conventionalist philosophy of mathematics. On the negative side, Carnap's reliance on infinitary methods becomes much more problematic than

[2] As Warren has recently shown, however, the reports of conventionalism's demise are greatly exaggerated (Warren 2020).

[3] Some of the key works are Friedman 1999a, Goldfarb and Ricketts 1992, Richardson 1994, Ricketts 1994, Friedman 1999b, Ebbs 1997, Awodey and Carus 2004, and Lavers 2008.

he took them to be. But, on the positive side, my interpretation gives a better explanation of the nature and importance of Carnapian analyticity than competing readings.

2 Carnap and Coventionalism

2.1 Mathematics as Analytic

Introducing Carnap's views on mathematics in a concise and unbiased way is difficult. For one thing, the exact nature of his position is precisely what we want clarity on. For another, due to his unusual metaphilosophy, Carnap's position cannot easily be compared to more mainstream approaches such as logicism, formalism, or Platonism. I will proceed by first outlining what Carnap says about the *analyticity* of mathematics, while remaining neutral about the philosophical significance this notion has. Answering the latter question will then occupy us for the rest of the paper.

Carnap's most important treatment of mathematics is his *Logical Syntax of Language* (Carnap 1937a). The bulk of this book does not contain philosophical discussions, but rather consists of a description of two formal theories for doing mathematics in: 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. For both of these theories, Carnap wants to draw a sharp distinction between analytic and synthetic sentences, with all of mathematics falling on the analytic side (Carnap 1937a: 116). In order to achieve this, Carnap thus needs to find the following:

> Analyticity
> A definition of 'analytic' for Language I and II such that: each sentence that only contains mathematical vocabulary is classified as either 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 the theorems of PRA analytic. 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 1937a: 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* are the usual recursive rules used in proof systems, such as modus ponens. For rules of this kind there is an effective procedure to determine whether they have been applied correctly. Rules of *consequence*, on the other hand, are not recursive. One example of the latter is the following infinitary rule:

ω-Rule

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

Adding the 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 1937a: 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 (yet), his definition of 'analytic' for Language II is effectively a version of Tarski's definition of truth (Carnap 1942: 247, Coffa 1987, Koellner ms). While we cannot go through all the details here, 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 1937a: 106)

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. In an earlier draft of *Logical Syntax*, Carnap had still tried to adhere to a purely syntactic approach in which quantification is interpreted substitutionally. But while this does not make much of a difference for first-order quantification, Gödel in correspondence pointed out that the approach fails for the higher-order quantifiers of Language II (Goldfarb 2003, Awodey and Carus 2010, Flocke 2019). In response to this problem, Carnap revised his account of quantification and added the following notion:

> By a possible *valuation* [...] we shall here understand a class [...] of accented expressions. (Carnap 1937a: 107)

Second-order quantification is then construed as quantification over all sets of accented expressions, where the latter are isomorphic to the natural numbers. The notion of a valuation thus arguably qualifies as a semantic concept, even though it took Carnap some more years to officially adopt this terminology (Ricketts 1996).

Without departing too much from the spirit of the original presentation, we can express the two definitions of analyticity as follows:

ϕ is analytic in Language I **iff** PRA + ω-rule $\vdash \phi$

ϕ is analytic in Language II **iff** $\mathbb{N} \vDash \phi$

So far this summary of what Carnap does in *Logical Syntax* should be uncontentious. The following questions then naturally arise: Why is the notion of analyticity *important*? Why, if at all, does it matter whether we classify mathematics in the way Carnap recommends? Answering them is difficult, and will directly lead us to the contested issue of conventionalism.

2.2 Metaphysical Conventionalism

Let us begin with the core idea behind linguistic conventionalism, which is succinctly summed up by Ayer. 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)

This formulation brings out what conventionalism is opposed to. Whereas Platonists hold that mathematical theories are about an independently existing realm of abstract objects, conventionalists deny that the model of language corresponding to entities provides an apt picture of the situation. And we do find Carnap saying analogous things about analyticity. He for instance writes that an analytic sentence "does not state anything about facts", since synthetic sentences concerning the observable world are the only "genuine statements about reality"

(Carnap 1937a: 41). Arithmetical claims ”do not concern experience or facts at all” (Carnap 1939: 56), because mathematical theories ”are systems of auxiliary statements without objects and without content” (Carnap 1953: 128).

Conventionalists also suggest a positive alternative to the Platonist account of mathematics. The key idea is that linguistic conventions we can adopt determine the meaning of mathematical expression, which are in turn said to entail all mathematical truths. This, again, lines up with what Carnap says about analyticity. The ”analytic character of a sentence depends solely upon the rules of application of the words concerned” (Carnap 1937a: 102), and thus we answer questions about mathematics through ”logical analysis based on the rules for the new expressions” (Carnap 1956a: 209).

In light of these remarks, it is thus not very surprising that many philosophers have understood Carnap's position as a form of linguistic conventionalism. And indeed, as I will argue, this thought is basically correct. But care is needed. The intuitive idea behind conventionalism can be spelled out in different ways, and some of the proposals that have been put forward are alien to Carnap's approach.

In an influential article, Boghossian describes a metaphysical conception of analyticity according to which a sentence S is metaphysically analytic if ”in some appropriate sense, our meaning p by S *makes it the case that p*” (Boghossian 1996: 365). Applied to the case of mathematics, the underlying picture appears to be this: Reality consists of facts. Some of the facts are about language use, others are about mathematics. Facts are connected in certain ways. Some depend on others, more fundamental facts are the grounds of derivative facts, and so on. The conventionalist thesis is thus a claim about the relationship between facts about language use and mathematical facts:

> Metaphysical Conventionalism
> Facts about the conventions concerning the use of mathematical language make it the case that mathematical facts obtain.

Boghossian primarily introduces the metaphysical conception of analyticity in order to dismiss it as untenable. And we can agree that, understood in the way described, conventionalism is not an attractive view. Among other things, it makes the truth of 2+2=4 *contingent*, since we might have used mathematical language differently than we actually do.[4] We thus need to ask: Do Carnap's remarks on analyticity commit him to metaphysical conventionalism?

[4] It is clear that Carnap, like Boghossian, takes this to be an undesirable result (Carnap 1937b: 37f).

The answer is no. New wave scholars, in particular Goldfarb and Ricketts, have convincingly argued that no conventionalist thesis relying on metaphysical notions like *facts* or *truthmaking* can be attributed to Carnap (Goldfarb and Ricketts 1992: 64).[5] Carnap does not conceive of reality as consisting of "language-transcendent" facts that stand in relations of dependence to each other. He has no room for "a notion of something's making a statement true" (Ricketts 1994: 177), there is no "thick notion of truth-in-virtue-of" (Ricketts 2007: 211). At most, linguistic conventions can be said to make it the case that analytic sentences *express* propositions that are guaranteed to be true. But there is no further claim about the conventions being what guarantees their truth.[6] Carnap's distinction between analytic and synthetic statements should thus not be understood as a metaphysical distinction between different truthmakers for the relevant facts. He is not a metaphysical conventionalist.

Should we also draw the stronger conclusion that there is *no* connection between analyticity and linguistic conventionalism at all? I don't think so. Carnap is not Quine. He insists on drawing a sharp distinction between the analytic and the synthetic, and it stands to reason that there must be *some* philosophical motivation for it. This view is also compatible with Goldfarb and Rickett's rejection of metaphysical conventionalism, for they are happy to write the following:

> 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)

What does the "flowing" of mathematical truth from the adoption of a metalanguage amount to? One might dismiss this as mere metaphorical fluff. But I don't think we should. In the next section, I will introduce a way to spell out this idea without using metaphysical notions alien to Carnap. After that I will then argue that we should actually ascribe a non-metaphysical form of conventionalism to him.

2.3 Commitment Conventionalism

In a reply to Quine, Carnap draws a distinction that is essential for our investigation:

[5] Ebbs 2001 is another important discussion of this point.

[6] This is inspired by Gillian Russell's defence of truth in virtue of meaning (Russell 2008), see Marschall 2022: 7-8 for the application to Carnap.

> [...] I should make a distinction between two kinds of readjustment in the case of a conflict with experience, namely, between a change in the language, and a mere change in or addition of, a truth-value ascribed to an indeterminate statement, [...]. A change of the first kind constitutes a radical alteration [...]. On the other hand, changes of the second kind occur every minute. (Carnap 1963: 921)

The distinction between change of language and change of truth value is connected to the analytic/synthetic distinction in the following way. The analytic sentences are determined by the rules of a language. If someone speaks a language in which a particular sentence is analytic, and later comes to reject this sentence, then – so Carnap – this entails that they have switched to speaking a different language. For synthetic sentences the situation is different. We can change our opinions about their truth values while remaining within one language.

It is helpful to spell out Carnap's distinction using the notion of *commitment*. The adoption of a language commits its users to accept certain rules of use, and those entail certain claims: namely the analytic ones. Consider a language that includes classical logic. A user of this language should clearly accept all instances of the law of excluded middle, since they are derivable from the classical inference rules. If a speaker persistently rejects the law of excluded middle, then it is natural to conclude that they do not accept the language with its classical rules at all, but rather one with a different logic. Synthetic sentences behave differently. While a language may provide rules of use for empirical concepts such as *rain*, it will usually not commit its users to accept particular claims about whether it is raining or not. Disagreeing with someone about empirical facts is thus not an indication that we speak different languages – at least if we are in broad agreement about what evidence speaks for and against the relevant claims.

More will be said about the notion of commitment and its relation to analyticity soon. For now I will use it to *formulate* the view I eventually want to ascribe to Carnap. I call it *commitment conventionalism* and will characterise it in two steps. The first component is the following thesis:

> There is a language L such that: every user of L is committed to accept every purely mathematical sentence as either analytic or contradictory.

This cashes out Goldfarb and Ricketts' metaphor. Mathematical truths "flow" from the adoption of L in the sense that its users are committed to accept them without further empirical investigations.

The second component is a thesis about this language L. There are two distinct flavours of commitment conventionalism that differ on this point. The first flavour, which I call *descriptive* commitment conventionalism, holds:

> The language we *actually* speak – for instance English or German – is like L.

In other words, the descriptive commitment conventionalist maintains that our current usage of English or another natural language commits us to accept or reject every purely mathematical sentence.

The second flavour of commitment conventionalism does not include this claim about natural languages, though it is in principle compatible with it. Instead, what I call *normative* commitment conventionalism makes a certain recommendation:

> At least for certain purposes, we *should* speak a language that is like L.

In other words, the normative commitment conventionalist proposes that, for specific purposes, we adopt a language which commits us to accept or reject every purely mathematical sentence.

According to the argument to come, Carnap accepts the normative version of commitment conventionalism. A descriptive version has recently been defended by Jared Warren (Warren 2020). In many ways Warren's account is congenial to the story I want to tell. He also rejects metaphysical conceptions of conventionalism (Warren 2020: 16f). I share his misgivings about severing all connection between Carnap and conventionalism. And, as we will see in section 5.1, his defence of descriptive conventionalism also helps the normative commitment conventionalist. However, I part company with Warren's claim that, for Carnap to be a conventionalist, he would need to maintain that conventionalism is "the uniquely true and correct theory of both logic and mathematics" (Warren 2020: 334). In particular, I object to the connection between conventionalism and what – following Coffa 1991: 322 – Warren calls *factualism* that is captured by the following quote:

> If Carnap thought that mathematical truths in natural languages were analytic, then he was a conventionalist. I suspect that he did, so I

> suspect that he was. But Carnap's ambivalence toward natural languages, as opposed to formal languages, further muddies the waters. (Warren 2020: 336, see also 329)

I am happy to admit that there is *some* evidence that speaks in favour of reading Carnap as holding that mathematical truths in natural language are analytic, which would make him a descriptive commitment conventionalist.[7] But, as we will see in the next section, it is clear that his primary goal is to defend the normative version of commitment conventionalism, not the descriptive one. And, contra Warren, I see no reason to maintain that only the descriptive view deserves the label conventionalism. After all, both flavours of commitment conventionalism draw on conventionally adopted linguistic rules to illuminate mathematical truth.

So much for a rough clarification of the target claim I plan to defend. We will now move on to the first step of my argument.

3 Step I: The Adoption Argument

3.1 Tolerance and Frameworks

Every discussion of the mature Carnap's position needs to take into account his unusual metaphilosophy. In what follows I will first argue that Carnap's famous principle of tolerance requires him to make room for the notion of commitment sketched above. Secondly, we will see that, according to Carnap's method of explication, philosophers are primarily in the business of making proposals rather than giving descriptions. Thirdly, I will combine these two results to argue that Carnap is indeed a normative commitment conventionalist.

The choice between different logics was lively debated at the time Carnap was writing *Logical Syntax*. Frege had regarded classical logic as enshrining the most general rules of thoughts. But competitors had emerged, such as intuitionistic logic. Which of these, if any, was the correct logic? With the adoption of the principle of tolerance, Carnap drew a radical conclusion: there is no such thing as a *correct* logic. There is no need to give a philosophical justification of a certain logic. The decision whether to use it rather has to be based on pragmatic considerations:

7 Carnap 1955, which we will discuss at the beginning of section 4, provides support for such a reading. See Goldfarb 1997 for a critical discussion of Coffa's notion of factualism.

> *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)

Carnap later introduced the label ″linguistic frameworks″ for what he here calls ″forms of language″ – i.e. languages together with certain rules of use and axioms. The idea that the adoption of such frameworks does not need to be justified deflates many philosophical debates, such as nominalistic concerns about quantifying over mathematical objects (Carnap 1956a). Rather than giving arguments for or against certain theses, the Carnapian philosopher becomes an engineer of sorts, who develops linguistic frameworks in the hope that they will be useful for the purposes of science and other practical pursuits.

As I will argue now, there is a close connection between Carnapian frameworks and the idea of commitment I relied on earlier. To get there it will be useful to re-formulate Carnap's principle of tolerance in the following way: There are no norms of correctness for the choice of frameworks. The adoption of a framework cannot be correct or incorrect in an objective sense, it can merely be fruitful or inadvisable given some aim one wants to achieve with it. But, importantly, Carnap does not hold the obviously untenable view that there are *no* norms of correctness governing any area of discourse at all. Instead, it is precisely the role of linguistic frameworks to *provide* such norms (Richardson 2007: 300, Ricketts 2007: 206, Steinberger 2017: 156).

Consider a framework which provides rules of evidence for talking about the physical world. Suppose that one of these rules states that smoke is evidence for fire. Further suppose that someone perceives smoke. Should this person consider it more likely that there is a fire around than they previously did? If we take the person to be committed to the rules of the aforementioned framework, then the answer should be yes. For if the rules of frameworks had no such effect on the way people ought to reason, then they seem wholly idle. But, on the the other hand, we should not be committed to the rules of *every possible* linguistic framework – there are many of those, and some are bound to prescribe conflicting responses to the perception of smoke. So in virtue of what can a person be bound by the rules of one framework but not to those of another?

I propose the following answer: Speakers can *adopt* a linguistic framework, which gives rise to a commitment to reason in accord with its rules. What exactly is the adoption of a framework by a speaker? Some account of this process is needed, and ideally one that explains the normative notion of commitment in terms of non-normative ones. Unfortunately Carnap himself does not say very much about the relationship between linguistic frameworks and actual speakers. He tends to focus on the description of the frameworks themselves and there are only very few remarks on what their adoption amounts to.[8] What is clear from the little he says, however, is that Carnap's conception of language is essentially a behaviourist one, according to which

> [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).

In light of this, the adoption of a linguistic framework must consist in the change or formation of certain linguistic dispositions of the speaker. This also fits an example of the adoption of a framework for describing temperature Carnap himself gives. 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. And this use, in turn, commits them to certain norms of correctness, such as that it is incorrect to ascribe two distinct temperatures to one and the same object.

This proposal presupposes that there are *facts* of the matter about which linguistic framework a speaker uses, which has been challenged in the secondary literature. I will postpone the detailed defence of this assumption until section 4, however, and take it for granted that it is legitimate for the rest of this section. We will now look at another crucial feature of Carnap's metaphilosophy, namely the notion of explication.

3.2 Explication and Direction of Fit

The description of framework adoption given above relies on a correspondence between framework rules and the dispositions of speakers. This is an important

[8] 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.

idea, but one whose significance for Carnap's project is easily misunderstood. One might think that, in constructing linguistic frameworks, Carnap is attempting to track and capture linguistic dispositions that speakers have. On this interpretation, Carnap is doing something like formal semanticists whose aim is to capture how people actually speak in systematic ways. The direction of fit goes from speech behaviour to formal frameworks: the latter are designed and altered to correspond to the former.

What is distinctive of Carnap's metaphilosophy, however, is that the intended direction of fit is the reverse of that adopted by formal semanticists.[9] 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. Such frameworks are put forward as part of *explications*, i.e. attempts to clarify and improve concepts:

> 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. (Carnap 1950: 3)

Unlike in formal semantics, frameworks thus play a normative rather than a descriptive role, and the intended direction of fit is from framework to speech behaviour. If an explication is regarded as useful, then speakers *should* amend their linguistic dispositions to be in accord with the suggested rules.[10] It is essential to keep this in mind, since, as Carus rightly stresses, it makes a big difference for how one can criticise a Carnapian framework:

> 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)

This can be illustrated using a famous example: Quine's rejection of analyticity (Quine 1951). While I don't want to make any claims about how this objection was actually meant to work, let us consider a straightforward way to interpret it:[11]

9 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). Despite this, he was not particularly interested in trying to capture the implicit rules speakers of ordinary languages follow (Carnap 1963: 931, 941).

10 Richard Jeffrey therefore describes Carnap's position as "voluntaristic" (Jeffrey 1994).

11 For an overview of the Carnap-Quine debate see Creath 2007. Some important papers are Creath 1991, O'Grady 1999, George 2000, Gregory 2003, and Hylton 2021.

(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 behaviouristic 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.

On Quine's reading, the thesis that there is a distinction between analytic and synthetic sentences is thus understood as a descriptive claim about natural language. But, even if we grant that (2) is correct, it is irrelevant to the project Carnap is engaged in. In an unpublished response to "Two Dogmas of Empiricism" he writes 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)

Like Quine, Carnap therefore thinks that the analytic/synthetic distinction does not correspond to any antecedent usage of the terms. It can only be drawn within a framework by means of explicit rules. 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 a proposed framework. For this it does not matter 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 project of explication and can therefore be rejected.

Earlier I distinguished between a descriptive and a normative version of commitment conventionalism. It should by now be clear why, contra Warren, I don't want to ascribe the descriptive version to Carnap. What speaks against it is that his reply to Quine is so concessive. If Carnap wanted to defend a descriptive thesis about natural language, then Quine's point is a real challenge, and Carnap should reply by maintaining that we do in fact have a stable practice of using 'analytic' in English. The actual reply, namely that Carnap primarily cares about formalised languages with explicit rules, would be beside the point. The analyticity of mathematics should therefore not be understood as a descriptive thesis about how people actually use mathematical language, but rather as a proposed explication to adopt certain rules for certain purposes.

3.3 Deriving Normative Commitment Conventionalism

I am now in a position to argue that Carnap accepts normative commitment conventionalism. The first ingredient is the aforementioned idea that the adoption of a Carnapian framework gives rise to certain commitments for the user. We can spell this out more explicitly as follows:

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

The second ingredient are Carnap's Languages I and II. As we already saw, in *Logical Syntax* Carnap provides finely crafted definitions of analyticity for these languages, which have the effect of classifying each purely mathematical sentence as either analytic or contradictory. Languages I and II are Carnapian frameworks that, so we may suppose, can be used. Applying the commitment principle just outlined to them, we thus arrive at the first tenet of commitment conventionalism:

> There is a language L such that: every user of L is committed to accept every purely mathematical sentence as either analytic or contradictory.

Some further remarks on the notion of commitment are in order before moving on. The disjunctive formulation that one is committed to accept every purely mathematical sentence as *either* analytic *or* contradictory is ambiguous. To see this, consider the case of truth. Carnap's frameworks usually employ classical logic, and therefore the principle of bivalence holds in them. It thus follows from the rules of the framework that *every* sentence S, whether synthetic or analytic, is either true or false. Consequently, every user of such a framework will be committed to accepting it as either true or false.

Is this disjunctive conception of commitment the same that is in play in the formulation of commitment conventionalism? No, it is too weak. In the case of synthetic sentences, adopting a framework commits one to accepting each of them as either true or false, but the rules leave open *which* of these truth values it is. The commitment is thus merely disjunctive. For analytic sentences, however the situation is different. Here the rules determine not just *that* a sentence is analytic or contradictory, but also which of these two options it is. The commitment this gives rise to is thus disjunctive but not merely so. A user of the framework is committed to accept certain sentences as analytic, and others as contradictory, and not just as being either one. To illustrate: Someone who accepts the rules of classical logic is committed to accepting statements of the form "$p \rightarrow p$" as true. Of course they are also committed to the weaker, disjunctive claim "$(p \rightarrow p) \vee \neg(p \rightarrow p)$". But the latter kind of commitment also arises for sentences that are no logical truths, such as "$p \wedge q$". In such a case one is still committed to the disjunction "$(p \wedge q) \vee \neg(p \wedge q)$", whereas for logical truths one is in addition committed to one of the disjuncts.

The second tenet of normative commitment conventionalism was this:

> At least for certain purposes, we *should* speak a language that is like *L*.

Now, I take it to be obvious that Carnap thought that his Languages I and II should be used for some purpose, for otherwise he would not have written a book about them. But what is this purpose? According to Carnap's own mission statement in *Logical Syntax*, his aim is to replace traditional philosophy by the "logic of science", which he describes as the "logical analysis of the concepts and sentences of the sciences" (Carnap 1937a: xiii). Carnap then suggests that philosophers aiming to study and improve the language of science should use Language I or II for this purpose:

> The book itself makes an attempt to provide, in the form of an exact syntactical method, the necessary tools for working out the problems of the logic of science. This is done in the first place by the formulation of the syntax of two particularly important types of language which we shall call, respectively, 'Language I' and 'Language II'. (Carnap 1937a: xiii-xiv)

This consideration is still quite abstract, since it remains open what features of Languages I and II make them especially attractive for the project Carnap wants to pursue. But this is a question we will postpone until section 4.2.

I take it that sufficient evidence for ascribing both tenets of normative commitment conventionalism to Carnap has been provided. But my argument is not complete yet. As I already flagged, my conception of framework adoption assumes that there are facts of the matter about whether a particular speaker uses a certain framework or not. Otherwise it would make no sense to describe a speaker as being committed to the rules of a framework. And while this assumption may seem innocent at first, it requires further scrutiny.

4 Step II: The Case for Adoption Facts

4.1 Full and Partial Adoption

I need to defend the following:

> Full Adoption Facts
> The speech dispositions of a speaker can determine whether they use a particular framework or not.

Let me begin by stressing that the assumption that there are adoption facts is very natural. Without them, it is hard to see what it would mean to use a Carnapian framework and bind oneself to its rules. And if one cannot do that, what is the point of constructing such frameworks in the first place?

Furthermore, in "Meaning and Synonymy in Natural Language" – one of the rare instances where Carnap explicitly discusses the relation between frameworks and natural languages – he seems to presuppose that there are adoption facts. Contra Quine, Carnap there argues that there are empirical criteria for attributing intensional notions to speakers (Carnap 1955: 37, see also Carnap 1963: 920). Accordingly, for a long time assuming adoption facts was the norm in the secondary literature on Carnap. In 1982 Ricketts takes Carnap to need a *criterion of analyticity* for "attributing a linguistic framework to an investigator" (Ricketts 1982: 124), and still explicitly makes this assumption in 1994:

> Not only can the syntax of the language in principle be read off from "speech habits;" [...] the syntax of a calculus [also] determines habits that would make one a speaker of a language with that syntax. (Ricketts 1994: 188)

In 1997, however, Gary Ebbs mounted the first sustained attack on adoption facts (Ebbs 1997: section 57-59). And in a paper from 2003, Ricketts renounced his earlier position and now denies that Carnap believes in – or needs there to be – facts about framework adoption (Ricketts 2003).[12]

We already saw that Carnap distinguishes between two kinds of rules: rules of derivation which are recursive, and rules of consequence which aren't. An example of the former kind of rule is modus ponens, an example of the latter kind is the infinitary ω-rule. According to Ricketts' new view, a linguistic framework that includes rules of both kinds can be adopted only *partially*. While speech habits "pin down a formal system for the language, Carnap's L-derivability" (Ricketts 2003: 277n9) – i.e. the recursive rules of derivation – they "do not fix the L-rules for a calculus instantiated by a language" (Ricketts 2003: 262) – i.e. the non-recursive rules of consequence. This is because 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" (Ricketts 2003: 262).

Ricketts does not deny that there are *any* facts about framework adoption at all:

> For a speaker's habits to agree with a calculus, Carnap appears to require [...] that the speaker not be disposed to affirm any contravalid sentence nor to deny any valid sentence. (Ricketts 2003: 262)

But adoption is partial rather than full, since it does not settle the non-recursive rules of consequence:

> Partial Adoption Facts
> The speech dispositions of a speaker *constrain* which frameworks can

[12] Limitations of space prevent me from discussing Ebbs' arguments in more detail here. Let me nevertheless flag one important point of disagreement. When discussing the relationship between formal calculi and natural language, Carnap stresses the need for *coordinative definitions*, a notion introduced by Reichenbach in the context of geometry. According to the latter, coordinative definitions have two roles. First, they can turn purely formal axiomatised systems of geometry into empirical theories about physical space. Furthermore, Reichenbach endorses a form of *geometrical conventionalism* according which, for instance, there are no objective facts about whether objects uniformly change in length when being moved through space or not (Reichenbach 1958: section 4). Resolving this indeterminacy is the second role coordinative definitions play. Ebbs interprets Carnap's assertion that the coordination of formal and natural languages is "perfectly analogous" (Carnap 1942: 12) to the case of geometry as committing him to *conventionalism about framework ascription*, i.e. the view that facts about speech dispositions leave open which framework a speaker uses. This indeterminacy is then resolved by coordinative definitions, which are akin to Quine's "analytic hypotheses". I think, however, that the analogy Carnap has in mind only concerns the first use of coordinative definitions, namely their role in giving formal calculi an empirical interpretation. Among other things, this is supported by the fact that Carnap's actual reaction to Quine's arguments for the indeterminacy of meaning (Carnap 1955: 27) is pretty much the opposite of what Ebbs predicts (Ebbs 1997: 120).

> be ascribed to them by fixing the rules of derivation. But they do not determine rules of consequence and therefore do not single out one particular framework.

If this is correct, then Carnap cannot be a commitment conventionalist. Commitment conventionalism requires that, for some language, any speaker of that language is committed to accept *all* purely mathematical sentences as either analytic or contradictory. This is only possible if speakers can commit themselves to the non-recursive rules of consequence of such a language, since they are needed to overcome the problem of incompleteness. On Ricketts' view, however, speakers cannot be said to *use* such rules of consequence at all. They can thus at most be committed to accept mathematical sentences that are entailed by the rules of derivation, which leaves the status of undecidable mathematical sentences unsettled.

In the following I will argue against Ricketts' view from "Languages and Calculi". I think that his earlier position, according to which Carnap does need a criterion of analyticity, is the superior interpretation. My argumentative strategy is as follows: If we accept full adoption facts and read Carnap as a commitment conventionalist, then we can easily explain why he cared about the analyticity of mathematics. If we follow Ricketts (and Ebbs) and reject these two things, then – so my claim – no compelling reason for the analyticity of mathematics can be given. Since Carnap clearly thought it important to classify mathematics in this way, my reading makes better sense of his position and is thus preferable.

4.2 Why Analyticity?

In *Logical Syntax*, Carnap notes a difference between his Languages I and II and the traditional systems put forward by, for instance, Frege and Russell. In the latter some sentences that only contain logical and mathematical vocabulary turn out to be descriptive (Carnap 1937a: 231). In particular, this is the case for sentences that are independent of the rules of derivation of the relevant systems. Carnap adds rules of consequence to his Languages I and II precisely to avoid this consequence. He wants *all* purely mathematical sentences to be classified as either analytic or contradictory, and hence as logical rather than descriptive. Any interpretation of Carnap's philosophy of mathematics thus needs to answer the following question: Why, for Carnap, is it important to classify mathematical sentences in this way? What, if anything, would be wrong with classifying undecidable mathematical sentences as descriptive?

On a conventionalist reading, these questions are easily answered. It was and still is commonly thought that there is a conflict between empiricism and the acceptance of abstract objects. Carnap rejected this alleged incompatibility, and argued that quantifying over mathematical objects "does not imply embracing a Platonic ontology but is perfectly compatible with empiricism and strictly scientific thinking" (Carnap 1956a: 242). This is possible – at least according to my reading – because mathematical claims are analytic, and thus follow from the rules of a linguistic framework. A user of such a framework is thus committed to accepting mathematical claims in virtue of them accepting certain rules. There is no need for any additional "Platonic ontology".[13]

If this strategy is to succeed, however, it should better cover *all* mathematical sentences. If some of them were still classified as synthetic and hence descriptive, as is the case in the systems of Frege and Russell, then the aforementioned worries about supposedly problematic ontological commitments can be raised concerning these sentences. In light of this, it is therefore not surprising that Carnap sees the need to use rules of consequence to give a definition of analyticity that is complete with respect to mathematical sentences.[14]

Ricketts needs a different explanation for Carnap's insistence on the analyticity of mathematics. In particular, he needs to address the following question: Since speakers cannot be said to actually follow rules of consequence, we can decide whether we want to describe sentences that are independent of the rules of derivation they use as analytic or synthetic. This decision has to be "guided by considerations of expedience" (Ricketts 2003: 273). So what makes classifying undecidable mathematical sentences as analytic expedient?

Earlier we saw that Carnap intends his Languages I and II to be used for reconstructing the logic of science. Drawing on the distinction between formal and factual science (Carnap 1953), Ricketts considers a concrete example of what this project involves. One of Carnap's aims is to capture the relationship between observable evidence and scientific theory using logical means. This is achieved by construing relations of evidential support as logical entailments. To use a simplistic example, the fact that smoke is evidence for fire might be expressed by means of the following conditional:

13 In this vein, Carnap for instance argues that he is able to accept Russell's axiom of infinity, which had been controversial in light of its ontological commitments, precisely because he interprets it as analytic (Carnap 1937a: 140f, Carnap 1956b: 86fn9, Lavers 2016).

14 In addition, two so far unpublished letters from Carnap to Irving M. Copi suggest that the former wanted to account for the a priori knowability of undecidable mathematical sentences by means of the rules of consequence (Marschall ms).

(C1) Fire $\rightarrow$ Smoke

The absence of smoke then logically entails that there is no fire present ($\neg$Smoke $\rightarrow$ $\neg$Fire). Given such a reconstruction questions about what evidence there is for and against fire make good sense, since they can be answered by pointing at the logical relations between statements enshrined by conditionals like (C1). Furthermore, there can be evidence against (C1) itself. It is refuted by the occurrence of fire without smoke. This logical relationship can be expressed as follows:

(C2) (Fire $\rightarrow$ Smoke) $\rightarrow$ $\neg$(Fire $\wedge$ $\neg$Smoke)

Statements about fire and smoke, as well as the conditional (C1) stating that there is a connection between them, are empirical claims about the world. Carnap thus wants to classify them as factual and synthetic. But what about the conditional (C2), which expresses a general logical fact about conditionals, negation, and conjunction? Is there evidence for or against it? Carnap thinks not. (C2) is merely an artefact of the acceptance of certain inference rules, which allow one to move from "(Fire $\rightarrow$ Smoke)" to "$\neg$(Fire $\wedge$ $\neg$Smoke)". In principle one could construct the language of science in such a way that (C2) cannot be expressed at all (Carnap 1953: 126). But this is cumbersome, and so it is more convenient to permit sentences like (C2). They are no factual claims, however, but a merely formal "auxiliary statements", "mere calculational devices [...] that [...] can be subjected to the same rules as [...] the genuine (synthetic) statements" (Carnap 1953: 126). For this reason, logical truths like (C2) are classified as analytic rather than synthetic.

Synthetic sentences are thus testable in virtue of being connected to other sentences, whereas analytic sentences express that such connections obtain, but are not themselves testable in the same way. According to Ricketts, making this distinction explicit is precisely why Carnap needs the notion of analyticity:

> Any calculus/semantic system that is a candidate to be the language for existing natural science must link theories to observation sentences. [...] In expedient languages, these links [...] generate L-true [analytic] and L-false [contradictory] sentences in the language. Consequently, these L-determinate sentences [...] are neither confirmed nor disconfirmed by the predictive success or failure of theories stated in the language. (Ricketts 2003: 271)

I think that this is a very compelling account of why Carnap wants to classify logic conceived in a *narrow* sense – i.e. propositional and quantificational logic – as analytic. It also makes sense of why Carnap regards the notion of analyticity as "indispensable" for science: there is no way to dispense with logical connections between sentences (Carnap 1963: 922, 932, Ricketts 2003: 278n18). So far so good.

The crucial question for our purposes, however, is whether this account can be generalised beyond logic to mathematics, and in particular to those parts of mathematics that are undecidable. Ricketts suggests that this generalisation can also be motivated by considerations of expedience (Ricketts 2003: 272). But, taken in the abstract, it is hardly obvious what this alleged expedience consists in. And I ultimately think that the generalisation fails. In order to make this case, I will discuss a particular type of undecidable sentences: those that express claims about consistency.

4.3 Consistency Sentences: A Case Study

The example I will rely on has been given by E. W. Beth in a critical discussion of *Logical Syntax* (Beth 1963). He imagines two logicians who both read Carnap's book but understand it in different ways. In Beth's article they are called Carnap and Carnap*, but since this terminology is apt to confuse I will name them *Yay* and *Nay*. Their disagreement concerns a sentence that, by Gödel's second incompleteness theorem, is independent of the rules of derivation of Language II: namely the consistency sentence Con_{II}. According to the rules of consequence Carnap puts forward, Con_{II} is analytic despite not being derivable (Carnap 1937a: 133). And, in line with this, Yay regards Con_{II} as true. Nay, on the other hand, rejects Con_{II} and regards its negation $\neg\text{Con}_{II}$ as true (Beth 1963: 478).

What are we to make of this scenario?[15] One crucial question is whether Yay and Nay are genuinely *disagreeing* about one and the same claim, or whether they are merely *appearing* to disagree but are really talking past each other. On the conventionalist reading of Carnap I endorse, the latter option is more natural. The following diagnosis suggests itself:

The Analytic Description

Yay and Nay are using different linguistic frameworks. In Yay's

[15] What exactly Beth himself wanted to achieve by presenting this scenario has been controversial, see Marschall 2021 for an (opinionated) discussion.

> framework Con_{II} is analytic in virtue of its rules of consequence. Nay's framework has the same rules of derivation as Yay's, but different rules of consequence. The latter make $\neg\mathrm{Con}_{II}$ analytic for Nay.

This, however, requires that frameworks can be adopted fully – including their rules of consequence – and not only partially. Since Ricketts rejects this assumption, he cannot say that the above description is *factually* correct. This is not to say that he cannot still recommend the analytic description. But one would have to give *pragmatic* reasons for preferring it over the following alternative:

> The Synthetic Description
> Yay and Nay are using the same linguistic framework, one whose rules leave the status of Con_{II} open. They are thus genuinely disagreeing about a synthetic sentence, just like on might disagree about whether it is raining.

Is this option less expedient than the analytic description? I don't see how. In fact, there seem to be strong reasons for favouring the synthetic description. Prima facie, it is not compelling to treat sentences like Con_{II} as purely formal auxiliary statements that are empty of content. Since Con_{II} can be taken to express the claim that Language II is consistent, i.e. that no contradiction can be derived from its rules of derivation, it is more natural to instead classify it as a factual statement. As Beth himself points out, this way of looking at things explains why the position taken by Nay will strike one as peculiar:

> Now [Nay] could settle the dispute at once in his favor by exhibiting the inconsistency which, according to him, exists in Language II, that is, by actually deriving a contradiction. But this he is unable to perform. (Beth 1963: 481)

Insisting that Language II is inconsistent without being able to actually demonstrate this is an eccentric view that borders on the unreasonable – Beth even calls Nay "psychopathic" (Beth 1963: 484). One can imagine plenty of evidence that supports Yay's position over Nay's, such as the fact that no contradiction has actually been found. All of these considerations favour the synthetic over the analytic description.[16]

Additional doubts about Ricketts' account are raised by Carnap's own reaction to Beth's example:

[16] For similar reasons Gödel objects to the idea that mathematics is empty of content in his critique of Carnap (Gödel 1995: 339).

> [...] it seems to me misleading to say that [Nay] 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) [Nay] 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 [...] (Carnap 1963: 929)

What distinguishes *ML* from *ML** is precisely that, in the former, Con_{II} is analytic, whereas in the latter $\neg\text{Con}_{II}$ is. In other words, Carnap unambiguously endorses the analytic description of the case, according to which Yay and Nay use different languages. But, importantly, he does not describe this description as preferably or expedient. Instead, he calls it more *correct*. This suggests that Carnap considered the question of which of the two description applies to be a factual one. And this indicates, contra Ricketts, that he believed in full and not merely partial facts about framework adoption.

This ends my case for interpreting Carnap as a conventionalist. The dialectical situation we have found ourselves in is intricate, since the textual basis is slim and open to interpretation. No doubt further responses on behalf of Ricketts (and Ebbs) are possible. Nevertheless, I take myself to have shown that those who reject full adoption facts face an uphill battle. They need to explain why Carnap insisted on the analyticity of all of mathematics even though rules of consequence are not actually used by speakers. As the previous discussion demonstrates, this is no easy task. For now I therefore take the competing and traditional assumption that linguistic frameworks are fully adopted by speakers to be in much better shape.

5 Evaluating Carnap's Conventionalism

5.1 Infinitary Rule Following

In the preceding two stages, I argued that Carnap's philosophy of mathematics is a position called normative commitment conventionalism according to which

> There is a language *L* such that: every user of *L* is committed to accept every purely mathematical sentence as either analytic or contradictory.

> At least for certain purposes, we *should* speak a language that is like L.

Let us suppose that my interpretation of Carnap is correct. Given this, is his view a *good* view? Does it deserve more attention than it currently receives in systematic discussions of the philosophy of mathematics? Obviously answering these questions in a comprehensive way would require more space than we have left. I will therefore focus on the most problematic aspect of Canap's conventionalism: namely that it recommends us to do something that many have thought *impossible*.

To see what I have in mind, consider Carnap's Language I. It is supposed to be one example of a language that, when adopted, commits its user to the analyticity or contradictoriness of every purely mathematical sentence. Suppose that we now want to follow Carnap's advice and actually adopt Language I, which will give rise to the aforementioned commitment. How are we to go about this? Well, we need to start following the rules of Language I. This is straightforward enough for Language I's recursive rules of derivation. But what about rules of consequence, like the infinitary ω-rule? Carnap himself seems to have thought that there is no difference between using it and a rule like modus ponens:

> 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)[17]

In other words, Carnap's position appears to be that we are able to follow both finitary and infinitary rules. This assumption, however, is questionable. Even if it turns out to be correct, it is not as obvious as Carnap makes it out to be. Most philosophers of mathematics that have commented on the matter have sided with Tarski against Carnap, and hold that we can only genuinely use rules that are finitary (Field 1994, Raatikainen 2005, Smith 2013: 332, Button and Walsh 2018: chapter 7). A detailed case in favour of the ω-rule being followable has only very recently been made by Warren (Warren 2020, Warren 2021). So while

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

Carnap's conventionalism cannot be discounted as easily as some have thought, it nevertheless relies on an assumption that remains contentious.

We see here an important logical connection between the normative and the descriptive versions of commitment conventionalism. While the normative version doesn't entail that the descriptive version is true, it entails something weaker: namely that descriptive commitment conventionalism *could be* true. Normative commitment conventionalism in effect says that we should make descriptive commitment conventionalism true by adopting a language in which mathematics comes out as analytic. This recommendation would make no sense if it were *impossible* to use such a language. But according to the critics of infinitary rule-following, this is precisely the problem. Provided that we restrict our attention to human beings, no one can follow Carnap's non-recursive rules of consequence, and so descriptive commitment conventionalism cannot be made true. Warren's own defence of infinitary rule-following, even though intended to support his own brand of descriptive commitment conventionalism, is thus also relevant for the prospects of Carnap's normative version of the view.

So much for Language I. Does Language II fare any better? On first inspection it might appear so. 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. And there seems to be nothing infinitary about a Tarskian truth definition. Its key component is a finite number of clauses which specify the truth conditions for different types of sentences. Here's a 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.

Language II thus doesn't seem to require its users to follow infinitary rules. Furthermore, the strategy of defining analyticity for it is more general than the route taken for Language I, and it is also the mature Carnap's preferred approach (Carnap 1963: 901). Can we thus disregard the concerns about infinitary rules raised above by only focussing on Language II?

Unfortunately the answer is no. For the semantic definition of analyticity gives rise to the following puzzle: How, given that a Tarskian truth definition can be given for *both* empirical *and* mathematical discourse, is mathematical truth still special? How, in other words, could it be that adopting Language II commits one to the analyticity of all true mathematical sentences, but entails no such commitment for synthetic sentences? A version of this worry was already raised by Quine. He claims that Carnap's use of semantic methods makes the thesis that mathematics is analytic trivial and uninteresting, since one could just as well classify the true sentences of physics in this manner:

> [Carnap's thesis says] 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)[18]

The good news: This challenge can be answered, and the next section will be devoted to spelling out how. But my proposal also comes with some bad news. The need for infinitary rules will reappear, and so the semantic strategy of Language II is not a way to dodge this issue after all.

5.2 Semantic Rules: A Proposal

Consider a formal theory for doing physics. Following Tarski's approach, we can give a definition of truth for the physical object language in a metalanguage. In this metalanguage we might then be able to state things like

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

[18] More recently Koellner has likewise pointed out that it is "quite unclear what it means to say that these [semantic] rules determine the truth value of a given statement" (Koellner ms: 34).

There is obviously no temptation to say that, in virtue of this biconditional, physical truth is somehow determined by, or flows from, the rules of the metalanguage. The clause only states the truth *conditions* of the statement '$\Psi(x, y, z, t)$'. Whether it is actually true or false is then determined by the properties of physical space, not by a linguistic rule.

As described earlier, Carnap's definition of analyticity for Language II – which is in effect a definition of mathematical truth – goes as follows:

ϕ is analytic in Language II **iff** $\mathbb{N} \models \phi$

Does this clause also merely state truth conditions of mathematical statements, like the biconditional for physical statements we just discussed? If so, then Quine's complaint that there is no interesting difference between truth in physics and mathematics would be vindicated. In order to overcome this challenge, it must be possible to understand the definition of analyticity as doing something stronger: namely to determine the truth *values* of mathematical claims, rather than just their truth conditions. And in order to see that such a reading is indeed possible, we need to scrutinise the role of *domains of quantification*.

The specification of a domain for the quantifiers to range over is a key part of Language II's definition of analyticity. Carnap's intended domain contains all and only the natural numbers, since he wants to capture the notion of truth in the standard model of arithmetic. We thus need to ask: what kind of activity is the specification of a domain for the quantifiers? Carnap explicitly writes that it is achieved by means of a linguistic rule:

> [...] in a *rule of values* related to the rules of designation, it is stated for each kind of variable which entities are to be *values* of the variables of that kind. Their class is sometimes called the *range of values* of the variables in question. (Carnap 1942: 44, see also Frost-Arnold 2013: 76)

But this does not suffice to clear up the matter, since the activity of "stating which entities are to be the range of variables" can be understood in different ways. Consider a simple case of talking about spatiotemporal objects: I assert that there is no more beer. It is plausible that, when making this assertion, I intend the relevant quantifier to range over the objects in a particular region of space, for instance all the things that are in my house. The assertion can thus come out true, whereas it will be false if the quantifiers were to range over all the things in the city I live in.

In this paradigmatic case of domain specification, two factors can be distinguished. First, I single out a certain part of the world – such as my house – that the quantifiers are supposed to range over. Secondly, the way this part of the world happens to be – what is contained in my house – then determines what objects are in the domain of quantification. Importantly, we can single out spatiotemporal parts of the world, such as my house, *by ostension*, without knowing or being acquainted with what objects this part of the world contains.

Things are very different, however, when it comes to abstract objects. Suppose that we want to specify the domain of quantification to be the standard model of arithmetic. Unlike houses, sets of mathematical objects are not be found in space and time. So how can we refer to a set that contains all and only the natural numbers in the specification of the domain? Since reference by ostension is not possible here, for Carnap the only other alternative is reference *by description*: One needs to single out the standard model of arithmetic by describing its content, i.e. the natural numbers.[19]

According to this proposal, the specification of a mathematical domain of quantification is thus an essentially different process from specifying an empirical domain. In the empirical case, specifying the domain by ostension does not entail that, through the act of specification, we know *which* objects the domain contains. As the beer example showed, this is determined by the way the world happens to be. 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.

It must be admitted that this interpretation goes beyond anything Carnap himself says. He does draw a distinction between descriptive and logical variables, but merely notes that "the whole question is in need of further study" (Carnap 1942: 59). Let me therefore stress one big advantage of my descriptivist proposal: It explains what is mistaken about Quine's complaint that Carnap's definition of analyticity is trivial and uninteresting. On my reading, to adopt Language II requires one to adopt the descriptive specification of the intended mathematical domain as well. In virtue of this, a user of Language II will then

[19] I am here following Button and Walsh's proposal for how one could "pick out particular isomorphism types" without "a special faculty of mathematical intuition" (Button and Walsh 2018: 151f). In fact, Carnap's project does not even require pinning down models up to isomorphism. It suffices that the models are *elementarily equivalent*, since they will then demarcate the analytic from the synthetic in the same way (Marschall 2021: section 5).

be committed to accept all sentences that are true in the standard model of arithmetic as analytic. And there is no analogous commitment for synthetic sentences, like those in physics. Their intended domain is specified by ostension, and so for them the rules of Language II only determine their truth value in tandem with the way the world is. Contra Quine, an essential difference between analytic and synthetic truth thus remains. This difference is a consequence of the different ways in which the domains of quantification for the respective areas of discourse are determined.

My reading, however, also comes with a downside. With regard to Language I, I expressed some reservations about whether the infinitary rule-following required by Carnap's conventionalism is actually possible. In a similar vein, there is now reason to doubt whether it is possible to describe the intended domain of quantification – the standard model of arithmetic – in a sufficiently determinate way. Indeed, the two points are basically the same worry. A full description of the standard model of arithmetic would have to entail what the truth values of all arithmetical statements are. And since there is no way to effectively enumerate all arithmetical truths, the need for non-recursive methods – plus the accompanying doubts – from the previous discussion recur.[20]

We thus end up with a positive and a negative conclusion. On the positive side, the adoption of Language II can be understood in a way that gives rise to the commitment required by Carnap's conventionalism. This rebuts Quine's triviality objection. On the negative side, the move to semantics does not circumvent the challenges posed by infinitary methods. Friends of Carnap's position thus need to devote their attention to the latter.

6 Conclusion

In this paper I did a number of things. First, I distinguished between three forms of conventionalism: metaphysical, descriptive commitment, and normative commitment conventionalism. I then argued that Carnap is a normative commitment conventionalist. In doing so, I defended the view that, for Car-

[20] In fact, the challenges posed by Language II might even be more severe, because it includes higher-order quantifiers. As is well known, the interpretation of second-order quantifiers is not fixed by specifying the domain of objects the first-order quantifiers range over. On a full semantics they range over the whole powerset of this domain, but alternative interpretations on which they only range over a subset are possible too. While I cannot investigate the repercussions of this fact for Carnap's position any further here, it would not be surprising if what is in effect conventionalism about set theory is even harder to defend than conventionalism about arithmetic.

nap, there are determinate facts concerning whether someone has adopted a certain language. Lastly, I evaluated the viability of Carnap's conventionalism by raising the challenge of infinitary rules. More could be said about each of these points. It would be presumptuous (and un-Carnapian) to claim that my interpretation of his philosophy of mathematics is the only defensible one. But, so I contend, from an exegetical standpoint my conventionalist reading is very attractive, since it explains many otherwise puzzling theoretical moves Carnap makes. In the end, the traditional view of Carnap as a conventionalist thus looks much better than recent scholarship has made it out to be.

Acknowledgements

I started writing this paper in order to address Hannes Leitgeb's detailed examiner comments on my PhD thesis *Carnap and the Ontology of Mathematics*. I thank him as well as Georg Schiemer, Tim Button, and Wouter Cohen, who have commented on previous drafts.

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.

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.

Ben-Menahem, Y. (2006): *Conventionalism: From Poincare to Quine*. Cambridge, England: Cambridge 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. A. (1996): 'Analyticity Reconsidered'. *Noûs*, 30(3):360–391.

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. (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. (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. (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. (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.

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.

Creath, R. (1991): 'Every Dogma Has Its Day'. *Erkenntnis*, 35(1):347–389.

Creath, R. (1992): 'Carnap's Conventionalism'. *Synthese*, 93(1-2):141–165.

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. (2001): 'Carnap's Logical Syntax'. In: Gaskin, R. (ed), *Grammar in Early Twentieth-Century Philosophy*, pp. 218–237. Routledge.

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.

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. (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. 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.

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.

Marschall, B. (2022): 'Carnap's Philosophy of Mathematics'. *Philosophy Compass*, 17(11):e12884.

Marschall, B. (ms): 'Carnap and the Epistemology of Mathematics'. Unpublished manuscript. Online: http://stirlingbus.com/bm.

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. (1936): 'Truth by Convention'. In: *Philosophical Essays for Alfred North Whitehead*, pp. 90–124. London: Longmans, Green & Co.

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.

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

Reichenbach, H. (1958): *The Philosophy of Space and Time*. Translated by Maria Reichenbach and John Freund. Dover Publications.

Richardson, A. (1994): 'Carnap's Principle of Tolerance'. *Proceedings of the Aristotelian Society*, 67:67–82.

Richardson, A. (2007): 'Carnapian Pragmatism'. In: Friedman, M. and Creath, R. (eds), *The Cambridge Companion to Carnap*, pp. 295–315. Cambridge University Press.

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.

Ricketts, T. G. (1982): 'Rationality, Translation, and Epistemology Naturalized'. *Journal of Philosophy*, 79(3):117–136.

Russell, G. (2008): *Truth in Virtue of Meaning*. Oxford University Press.

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

Steinberger, F. (2017): 'Frege and Carnap on the Normativity of Logic'. *Synthese*, 194(1):143–162.

von Neumann, J. (2005): *John Von Neumann: Selected Letters*. American Mathematical Soc.

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

Warren, J. (2021): 'Infinite Reasoning'. *Philosophy and Phenomenological Research*, 103(2):385–407.