

ARTICLE

# Carnap and Beth on the Limits of Tolerance

Benjamin Marschall<sup>1,2</sup>

<sup>1</sup>Darwin College, University of Cambridge, Cambridge, UK and <sup>2</sup>Institute Vienna Circle, University of Vienna, Vienna, Austria  
Email: bm515@cam.ac.uk

## Abstract

Rudolf Carnap's principle of tolerance states that there is no need to justify the adoption of a logic by philosophical means. Carnap uses the freedom provided by this principle in his philosophy of mathematics: he wants to capture the idea that mathematical truth is a matter of linguistic rules by relying on a strong metalanguage with infinitary inference rules. In this paper, I give a new interpretation of an argument by E. W. Beth, which shows that the principle of tolerance does not suffice to remove all obstacles to the employment of infinitary rules.

**Keywords:** Rudolf Carnap; E. W. Beth; principle of tolerance; philosophy of mathematics; linguistic conventionalism; incompleteness

## 1. Introduction

In his *Logical Syntax of Language*, Rudolf Carnap develops an account of the nature of logic and mathematics that differs radically from the views of his predecessors and contemporaries (1937a). At the heart of *Logical Syntax* is the *principle of tolerance*, according to which we can freely adopt any system of logic we like without further philosophical justification. This principle plays a crucial role in Carnap's philosophy of mathematics. Since it licences the use of *strong metalanguages* with infinitary inference rules, Carnap thinks that he is able to capture the idea that mathematical truth is determined by linguistic rules despite the limitative results of Gödel's incompleteness theorems. We thus end up with a view that seems attractive all around: classical logic and mathematics can be used, but there is no need to deal with the metaphysical questions associated with the acceptance of a Platonist ontology (Carnap 1956a). So what's not to like?

In this paper I argue that one of the most damaging objections to Carnap's philosophy of mathematics—namely E. W. Beth's *argument from nonstandard models* (Beth 1963)—has not received the attention it deserves so far. The upshot of Beth's paper is that there is a deep tension between two different roles Carnap has for metalanguages. On the one hand, Carnap needs a strong metalanguage to define mathematical truth, but, on the other hand, a metalanguage is also something that speakers *use* and *share*. The problem, so my reading of Beth, is that Carnap cannot account for the latter feature if metalanguages are sufficiently strong for the former task.

In the next section, I will outline Carnap's philosophy of mathematics with an emphasis on the role of tolerance in dealing with Gödel's incompleteness theorems. Section three contains a detailed reconstruction of Beth's argument from nonstandard models, including a discussion of Carnap's own dismissive reaction to it. In section four, I then show that the force of the argument has been underestimated since, rightly understood, Beth has correctly diagnosed a serious problem for Carnap's reliance on infinitary inference rules to overcome incompleteness. Section five concludes the paper by comparing my reading of Beth to other well-known model-theoretic arguments.

## 2. Carnap's philosophy of mathematics

### 2.a Tolerance, analyticity, and conventionalism

When it comes to logic and mathematics, Carnap's main influences were Frege, Russell, and the early Wittgenstein. Despite their many differences, these philosophers share a monistic conception of logic according to which there is *one true* system of logic. The development of intuitionist logic, for instance, was seen as a challenge to the orthodox view that (some form of) classical logic is the correct one, with philosophical disputes ensuing. Carnap, on the other hand, wants to dispense with skirmishes of this kind, as he stresses in the introduction to *Logical Syntax*:

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

Later in the book, Carnap sums up his own position in the form of a well-known set of slogans:

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

There has been plenty of discussion of the principle of tolerance already, and friends of Carnap's position have fended off various objections to it, such as the charge that it presupposes an untenable form of verificationism (Putnam 1983; Ricketts 1994; Putnam 1994). In this paper, I will focus on one particular use of the principle of tolerance, however—namely its role in Carnap's philosophy of mathematics.

Let us begin by sketching Carnap's position in *Logical Syntax*. In the book, he describes two languages which contain mathematical vocabulary and come with certain rules. Carnap then calls a sentence of such a language *analytic* if it follows from the rules the language, *contradictory* if its negation follows, and *synthetic* if it is independent of the rules. Ultimately, he wants to establish that in his preferred language for mathematics the following holds:

Every purely mathematical sentence, i.e. every sentence which only contains mathematical vocabulary, is either analytic or contradictory. (adapted from Carnap 1937a, 116)

This is supposed to capture the classical idea that every purely mathematical sentence is either determinately true or determinately false. Put differently, for mathematical sentences *truth* and *analyticity* (as well as *falsity* and *contradictoriness*) coincide. For this reason, Carnap thinks that the acceptance of classical mathematics does not commit us to anything metaphysically problematic, such as a Platonist ontology (compare Carnap 1937a, 114).

Since mathematical truth is supposed to be a consequence of the adoption of certain rules, the question of whether Carnap's position is a form of *linguistic conventionalism* arises. In recent years scholars such as Gary Ebbs (2011) have argued that, for instance, Quine's well known anticonventionalist arguments in “Truth by Convention” (1949) are not a problem for Carnap. This is because conventionalism is usually understood as the claim that the truth of mathematical statements can be *explained* in an informative and noncircular way in terms of conventions, but, as Thomas Ricketts (2007, 211) stresses, Carnap has no place for the relevant notion of explanation that is being invoked: “he rejects any thick notion of truth-in-virtue-of.”

Such readings of Carnap are called *deflationary*, since they stress his refusal to engage with many of the explanatory projects that were (and still are) part of the philosophical mainstream. While I think that the deflationary interpreters are largely right, it would be an exaggeration to claim that there is *no* similarity between Carnap's position and linguistic conventionalism at all. Warren Goldfarb and Ricketts themselves are perfectly happy to describe Carnap's position as one according to which "mathematical truths [...] flow from the adoption of the metalanguage" (1992, 71), for instance, and Ebbs also grants that there is an unproblematic sense in which, for Carnap, rules *determine* truth (2017, 26). The distinguishing mark of Carnap's brand of conventionalism is that he allows the rules from which mathematical truth is supposed to follow to be formulated in a *mathematical* metalanguage. Goldfarb describes the circular nature of Carnap's approach as follows:

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

This is in line with Carnap's acceptance of tolerance. Since there is, in general, no need to justify the acceptance of a certain language, we do not need to justify the acceptance of a mathematical metalanguage either. It would therefore be inappropriate to say that we can only use mathematical languages *because* mathematics is conventional. Carnap's position is thus an unusual one, but, so far, appears stable.

### 2.b Gödel and incompleteness

Carnap's tolerant attitude towards language choice also plays a crucial role in his response to what seems to be a decisive obstacle to *any* version of conventionalism: Gödel's incompleteness theorems. In their most general form, Gödel's theorems show that there is no theory *T* which has all of the following properties:

- (1) *T* is consistent.
- (2) *T* is recursively formalised.<sup>1</sup>
- (3) *T* is strong enough to do basic arithmetic.
- (4) *T* is complete, i.e. for any *T*-sentence  $\phi$ , either *T* proves  $\phi$  or *T* proves  $\neg\phi$ .

On the face of it, this result seems to be a problem for Carnap. After all, he wanted his definition of analyticity for the mathematical language to be *complete*: for every purely mathematical sentence, either it or its negation was supposed to follow from the rules of the language. It is thus natural to think that Gödel's theorems in themselves undermine Carnap's project.

Carnap was aware of Gödel's results, however, and thought that he had a way to overcome this apparent problem. In effect, his strategy is to adopt a very liberal conception of what rules are admissible, which allows the use of *nonrecursive* rules to specify mathematical truth. To illustrate this, consider the well-known  $\omega$ -rule:

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

<sup>1</sup>More specifically, condition (2) requires that there must be a way to enumerate the axioms of *T*, and to mechanically decide whether its inference rules have been applied correctly (Raatikainen 2020, sec. 1.1).

If we add the  $\omega$ -rule to an arithmetical theory such as Peano arithmetic (PA), we get a complete theory. This does not violate Gödel's incompleteness theorems, because PA+ $\omega$ -rule is not a recursively formalised theory. Since the  $\omega$ -rule has infinitely many premises, it is impossible to mechanically decide whether the rule has been correctly applied in a particular case.

To Gödel himself, it seemed obvious that the claim that mathematical truth is a matter of what follows from the rules of a language only makes sense as long as the rules rely on “finitary concepts referring to finite combinations of symbols”—this requirement, so Gödel, “should be beyond dispute” (1995, 341).<sup>2</sup> This was (and arguably still is) the standard view among philosophers of mathematics, but Carnap explicitly disagreed. After outlining the  $\omega$ -rule in *Logical Syntax*, he addresses concerns about its infinitary nature as follows:

Tarski discusses [... the  $\omega$ -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. (1937a, 173)

At the end of the day, I think Beth's argument shows Carnap's last claim to be mistaken. But for the moment, we can note that Carnap apparently thought that the principle of tolerance allows him to use infinitary rules in his philosophy of mathematics.<sup>3</sup> Before he had accepted tolerance Carnap seems to have shared the widespread skepticism towards the  $\omega$ -rule, and, in 1931, described Hilbert's employment of this rule as “highly questionable” (Carnap, diary entry on 30.8.1931, quoted in Buldt 2004, 242). None of this caution remains in *Logical Syntax*, however, and it likely that Carnap concluded that only philosophical prejudices had stopped him from using this rule so far.<sup>4</sup>

Once again, Carnap's approach is unorthodox, but appears internally consistent given the freedom the principle of tolerance provides. We will now move on to Beth's argument from nonstandard models, however, which challenges Carnap's reliance on tolerance in his philosophy of mathematics. The question of infinitary rules will then be revisited in section 4.

### 3. Beth's argument from nonstandard models

#### 3.a Beth's objection and its reception

In his contribution to the Schilpp volume on Carnap, Beth argues that the possibility of interpreting formal languages in a model-theoretically nonstandard way poses a problem for Carnap's philosophy. More specifically, Beth claims that the possibility of nonstandard interpretations shows the following:

BETH'S THESIS

Carnap needs to rely on an intuitive interpretation of the metalanguage in *The Logical Syntax of Language* that is not fully captured by the explicitly formulated rules. This entails that the

<sup>2</sup>These quotes are from Gödel's unpublished paper “Is Mathematics Syntax of Language?” (1995), which aims to refute Carnap's position by relying on the second incompleteness theorem. There are some similarities to Beth's argument, but overall Gödel focusses much more on *epistemic* issues, and hence I will not go further into the details here. For discussion see Goldfarb (1995), Potter (2000, chap. 11), Awodey and Carus (2004) and Lavers (2019).

<sup>3</sup>While Carnap was happy with infinitely many premises, he curiously objected to infinite chains of rule-application (1937a, 106).

<sup>4</sup>Carnap's attitude towards impredicative definitions underwent a similar development. In a paper that first appeared in 1931, Carnap still thought that impredicative definitions commit one to Platonist metaphysics (1983, 50). In *Logical Syntax* he came to hold that, rightly understood, impredicative definitions are perfectly acceptable and metaphysically neutral (1937a, 114). For an illuminating study of how this change in view came about see Flocke (2019).

principle of tolerance cannot be maintained without restrictions. (adapted from Beth 1963, 479f)<sup>5</sup>

Beth introduces the following distinction to make his point:

In a treatise such as *Logical Syntax*, natural language can be used in two different ways, which I should like to denote as *strict usage* and *amplified usage*, respectively. In strict usage of natural language, we refer to a definite model of the theory to which our statements belong; it is this model which has been called the *intuitive model*. In amplified usage of natural language—and in *all* usage of formalised languages—on the other hand, we refer to *any* model of this theory. (Beth 1963, 479f)

Let us illustrate this by considering a formal theory such as first-order Peano arithmetic (PA). The intended model for the axioms of PA is one where the domain contains all and only the natural numbers 1, 2, 3, ... . This is called the *standard model*  $\mathbb{N}$  of arithmetic. It is a consequence of metalogical results such as the Löwenheim-Skolem theorem and Gödel's incompleteness theorems, however, that the axioms of PA do not uniquely pin down the standard model, or even a class of models isomorphic to it. While the standard model is countable, for instance, PA also has *uncountable* models, which are regarded as *nonstandard* interpretations.

In broad strokes, Beth's argument begins as follows:

- (1) What Carnap says in *Logical Syntax* must be read as involving strict usage of language, i.e., Carnap has one particular intended interpretation in mind.
- (2) This intended interpretation is not pinned down by the explicit statements Carnap makes in *Logical Syntax*: someone could read the book but interpret it in an unintended way.
- (3) This shows that there is a sense in which Carnap cannot replace all appeals to an intuitive notion of interpretation by explicit rules.

This part of Beth's paper has been discussed in the secondary literature, and commentators tend to agree with Beth's conclusions:

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

What becomes apparent in this and other representative passages (such as Goldfarb and Ricketts 1992, 72), however, is that Beth's objection has not been regarded as a deep challenge that goes to the heart of Carnap's view. Rather Beth has largely been read as flagging that Carnap overstates his case when he exclusively praises the virtues of exact rules, which is unfortunate, but more a case of misleading advertising than a challenge to the coherence of his position.<sup>6</sup>

<sup>5</sup>Furthermore, Beth maintains that the need for an intuitive interpretation also shows that Carnap cannot treat questions concerning ontology as a merely pragmatic matter (1963, 500f).

<sup>6</sup>One notable exception is Michael Friedman. In a paper that originally appeared in 1988, he argues against Carnap's strategy of overcoming incompleteness by means of nonrecursive inference rules (1999a), and explicitly draws the connection to Beth. This criticism spawned a debate throughout the 1990s (Goldfarb and Ricketts 1992; Richardson 1994; Friedman 1999b), which led Friedman to change his mind. In Friedman (2009, 240), he regards Carnap's own response to Beth as adequate. I think that the earlier Friedman was closer to the truth, but his interpretations of Beth are less effective than the one I will give.

In the following, I will argue that the real force of Beth's point has not been appreciated so far. For this Beth himself is partly to blame, however, since he did not present his argument in a particularly clear way. It will therefore be helpful to start with a detailed analysis of the text.

### 3.b Carnap and Carnap\*

Beth's core idea is that a logician whom he calls Carnap\* might interpret the language in which *Logical Syntax* is written in a nonstandard way. Our first task will be to understand what exactly Carnap\*'s nonstandard interpretation amounts to, as Beth's own remarks on this are sometimes confusing. The original presentation focusses on Carnap's language II, which is a version of the simple theory of types with higher-order quantifiers. In my reconstruction I will continue to talk about first-order PA, however, since nowadays this is a much more familiar theory. This change does no harm since, on my reading, the problem Beth identifies is not specific to a particular arithmetical theory.

Beth begins by pointing out that PA has nonstandard models in which some sentences receive different truth values from those they have in the standard model. This is illustrated by way of the consistency sentence  $\text{Con}_{PA}$ , which is not derivable from PA but true in the standard model of arithmetic. Calling the standard model  $M$ , Beth generates a nonstandard model  $M^*$  by adding  $\neg\text{Con}_{PA}$  to the axioms of PA. He then describes Carnap\* as a logician "whose logical and mathematical intuitions are in accordance with model  $M^*$ " (1963, 478).

At this point, it is tempting to understand the scenario as follows: Carnap and Carnap\* both look at the axioms of PA. Carnap interprets them as being about the natural numbers, while Carnap\* thinks that they should be interpreted with respect to the nonstandard model  $M^*$ . If they consider the truth of  $\text{Con}_{PA}$ , for instance, they will have different opinions.

As Carnap describes in his reply to Beth, however, this cannot be how the scenario is intended:

But now Beth proceeds to make a number of further statements about Carnap\* which at first glance appear as obviously false, e.g., the statement that the set of all axioms of  $\text{II}^*$  is different for Carnap\* and for us, i.e. Carnap and Beth, and the statement that for Carnap\* the languages  $\text{II}^*$  and II coincide. (1963, 929)

In our case, language II is PA and  $\text{II}^*$  is  $\text{PA} + \neg\text{Con}_{PA}$ . Clearly these are distinct theories since they have different axioms, so why would Beth say that for Carnap\* these languages coincide? Here is Carnap's take:

Beth's statements are understandable only on the basis of an additional assumption, namely that Carnap\* interprets not only the symbolic object languages but also the metalanguage  $ML$  in a way different from Carnap. Therefore I suppose that Beth makes this additional assumption, although he does not state it explicitly. (1963, 929)

This seems correct: for Beth's claims to make any sense, it must be that the metalanguage Carnap\* uses is nonstandard in some way. In particular, Carnap\*'s metalanguage  $ML^*$  must be such that, in it, Carnap\* can conclude that PA is inconsistent, for then it is clear why he cannot distinguish between PA and  $\text{PA} + \neg\text{Con}_{PA}$ : adding a sentence to an inconsistent theory just gives one the same inconsistent theory once again.<sup>7</sup>

How exactly does Carnap\* manage to prove the inconsistency of PA in his metalanguage though? As Beth describes it, when Carnap and Carnap\* both go through *Logical Syntax*, they will agree up to the following point:

<sup>7</sup>Some of Beth's remarks remain confusing. He, for instance, suggests that for Carnap\* languages II and  $\text{II}^*$  have no finite axiomatisation (1963, 478n27), which is false if Carnap\* regards these languages to be inconsistent. As I will discuss in section 5.b this is not the only incorrect or misleading statement Beth makes in his paper.

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

Furthermore, Beth writes that this difference in how to interpret the meaning of “for all syntactical properties of accented expressions” has the consequence that Carnap\* will reject Theorem 36.6 of *Logical Syntax*, which is just the statement that  $\text{Con}_{PA}$  is true even though not derivable.

The consistency sentence  $\text{Con}_{PA}$  is the following statement:

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

It is thus true if there is no number that encodes the proof of a contradiction from the axioms of PA. Now although  $\forall x \neg Pr_{PA}(x, \ulcorner \perp \urcorner)$  cannot be derived from PA, each instance of it *is* derivable. So PA entails all of the following statements:

$$\neg Pr_{PA}(0, \ulcorner \perp \urcorner)$$

$$\neg Pr_{PA}(0', \ulcorner \perp \urcorner)$$

$$\neg Pr_{PA}(0'', \ulcorner \perp \urcorner)$$

...

Carnap uses this fact to argue for the truth of  $\text{Con}_{PA}$  in his metalanguage. His definition of the analyticity of universal statements is as follows: in order to determine whether “ $\forall x P_1(x)$ ” is analytic, it is necessary 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. (1937a, 106)

And since Carnap assumes that the infinite sentential class at issue here consists solely of the standard numerals 0, 0', 0'', ... , this delivers the result that  $\forall x \neg Pr_{PA}(x, \ulcorner \perp \urcorner)$ —i.e.,  $\text{Con}_{PA}$  is true. Put differently, Carnap thinks that we can establish the truth of undecidable sentences such as  $\text{Con}_{PA}$  by using something like the  $\omega$ -rule in the metalanguage.

It is precisely this last step that Carnap\* takes issue with. He interprets the range of the quantifiers differently from Carnap, and admits some numerals as substitution instances that are nonstandard, i.e., they are not generated in a finite number of steps starting at “0.” And not only does Carnap\* believe that there are such nonstandard numerals, but he furthermore thinks that one of them encodes the proof of a contradiction from the axioms of PA, which is why he holds  $\text{Con}_{PA}$  to be *false*. The case of Carnap\* is thus similar to a scenario Jared Warren describes in a recent paper, where some Martians use a version of Peano arithmetic that contains the following additional inference rule (2015, 1360):

$$\frac{}{\neg \text{Con}_{PA}}$$

Beth himself happily admits that this is a strange—even “psychopathic” (1963, 484)—conviction to have, since Carnap\* is not able to actually produce a syntactic proof of a contradiction from the

axioms of PA (1963, 481). But what matters for now is that the case of Carnap\* is coherent and intelligible.

### 3.c What's the problem?

We have now described what Carnap\* is like in some detail. The crucial question then becomes: Is the possibility of Carnap\* a problem for Carnap? And if so, why? Let us begin by looking at how Beth himself describes the connection between nonstandard interpretations and the principle of tolerance:

It should be noted that we also meet here with a limitation regarding the Principle of Tolerance. Indeed, Carnap could be tolerant with respect to Carnap\*, for Carnap would be able to understand why Carnap\* adopting for certain personal reasons additional axioms for Language II is compelled to accept additional theorems and to reject certain (and indeed all) models for language II. But Carnap\* would never be able to understand why Carnap, having accepted certain axioms and certain rules of inference, as stated in *Logical Syntax*, stubbornly refuses to accept non- $W_{II}$  as a theorem and believes Language II to have a model. (1963, 479)

We already discussed what Beth is alluding to here: namely that while from Carnap's perspective PA, PA+Con<sub>PA</sub>, and PA+¬Con<sub>PA</sub> are all distinct theories, Carnap\* cannot draw the same distinctions, since for him they are all inconsistent. From this observation one can draw a more general lesson, namely that there is a sense in which the principle of tolerance *itself* can be interpreted nonstandardly. After all, Carnap demands that logicians "give syntactical rules instead of philosophical arguments," but there is no guarantee that everyone shares our conception of syntax, and hence there can be disagreements about what counts as a syntactical rule.

Some commentators, such as Friedman, think that this is the crucial upshot of Beth's argument:

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

One might wonder whether the fact that it is possible to interpret the principle of tolerance in an unintended way is really an *objection* to it though. Carnap himself did not think so, and he sounds noticeably unmoved by Beth's argument in his reply from the Schilpp volume:

Since the metalanguage *ML* serves as a means of communication [...] I always presupposed both in syntax and in semantics that a fixed interpretation of *ML*, which is shared by all participants, is given. This interpretation is usually not formulated explicitly; but since *ML* uses English words, it is assumed that these words are understood in their ordinary senses. The necessity of this presupposition of a common interpreted metalanguage seems to me obvious. (1963, 929)

Carnap's view on the matter seems to be that while someone who interprets syntax nonstandardly like Carnap\* is *possible*, this is no reason to worry, since it is entirely reasonable to assume that the readers of *Logical Syntax* use a metalanguage with a standard interpretation of syntax, i.e., something like *ML*.

This relaxed reply is not without appeal. For as Carnap points out, it is also very important that readers of *Logical Syntax* do not interpret the expression "no occurrence" as meaning "at least one occurrence" (1963, 929), but surely the possibility that someone *could* misinterpret the book in this

way does not amount to an objection. What matters is that we do not *actually* misinterpret each other in the way Beth envisages, but demanding that misinterpretation is *impossible* is asking for too much.

I think that this would be an effective reply, provided that Carnap is entitled to say that we *de facto* use *ML* rather than *ML\** as our metalanguage. As I will argue now, however, this presupposition is less innocent than Carnap suggests. Exegetical issues will be set aside for a while, but in [section 5](#) I will address the question of whether the argument I am about to present can plausibly be ascribed to Beth.

## 4. Tolerance and mathematics

### 4.a Metalanguages and how to use them

In his reply to Beth, Carnap regards it as obvious that we use metalanguage *ML*. But without closer inspection, it is not so clear what to make of this claim, since what we speak in the first instance is a natural language such as English or German and not a formal language with explicit rules such as *ML*. In order to assess Carnap's response, we thus need to understand what it means to *use* a formal system *as* the metalanguage.

One of the few passages in which Carnap discusses natural language is the following:

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. (1939, 3)

Languages understood in this broadly behaviouristic sense are obviously not the same thing as languages conceived of as systems of explicit rules. But these two notions of language are not completely unrelated either. Carnap thinks that we can *coordinate* languages understood as formal calculi with languages understood as systems of dispositions in the following way: a population of speakers can be said to use a calculus if the explicit rules of this calculus correspond to linguistic dispositions the speakers actually have (1939, 5f). If, to take a simple example, speakers are disposed to reason in accord with *modus ponens*, then it is appropriate to capture this fact using a calculus that contains *modus ponens* as an inference rule.<sup>8</sup>

The situation is complicated by the fact that our linguistic behaviour does not unambiguously determine one unique set of rules, as Carnap points out himself:

Suppose we have found that the word “mond” of B was used in 98 per cent of the cases for the moon and in 2 per cent for a certain lantern. Now it is a matter of our decision whether we construct the rules in such a way that both the moon and the lantern are designata of “mond” or only the moon. If we choose the first, the use of “mond” in those 2 per cent of cases was right —with respect to our rules; if we choose the second, it was wrong. (1939, 6)

What should we conclude from this underdetermination? At times, Carnap sounds like he is endorsing the radical thesis that there are *no* objective standards of correctness at all when it comes to coordinating formal calculi with speech behaviour (1939, 6f). But this attitude would be hard to square with the reply to Beth we saw in the previous section, since there Carnap does rely on the fact that our ways of speaking correspond to metalanguage *ML*. And Carnap's later writings indeed corroborate a reading on which he believes in facts about which rules correspond to the way natural

<sup>8</sup>I rely on Ricketts's view that, for Carnap, logical calculi need to be coordinated with the “logically amorphous” natural language (2003). This interpretation has been challenged by André Carus, who holds that instead of a sharp divide between natural and formal languages Carnap sees their relationship as *continuous* (Carus 2004; Carus 2007; for discussion, see Richardson 2012 and Wagner 2012). I will discuss how Beth's argument fares on Carus's reading of Carnap in future work.

language is used. In a response to Quine that can also be found in the Schilpp volume, Carnap for instance writes the following:

It seemed rather plausible to me from the beginning that there should be an empirical criterion for the concept of the meaning of a word or a phrase, in view of the fact that linguists traditionally determine empirically the meanings, meaning differences, and shifts of meanings of words, and that with respect to these determinations they reach a measure of agreement among themselves which is often considerably higher than that reached for results in most of the other fields of the social sciences. Quine's arguments to the effect that the lexicographers actually have no criterion for their determinations did not seem at all convincing to me. (1963, 920)

This is quite a strong claim: contra Quine, Carnap thinks that empirical investigations can tell us what the meanings of words and sentences of natural languages are, where this is taken to include both their extension *and* their intension. In light of this, the assumption that there are facts about which inference rules we follow is a relatively innocent one.

What Carnap thought is one thing, but whether his assumptions are *justified* is, of course, another matter. In light of Quine's later thesis of the indeterminacy of translation (1960), as well as Kripke's interpretation of Wittgenstein's remarks on rule-following (1982), it is natural to suspect that Carnap was overly optimistic here. The reply to Quine as well as the paper "Meaning and Synonymy in Natural Languages" (Carnap 1955) both suggest that Carnap took his claims about empirical criteria of meaning to be supported by his account of linguistic dispositions, but the explicit discussion is quite brief.

A full investigation of Carnap's treatment of dispositions will not be possible here, as it would require a more extensive survey of his views on scientific theories, laws of nature, and modality.<sup>9</sup> For the purposes of this paper, I will therefore make the dialectically generous assumption that Carnap is in fact justified to maintain that we have dispositions corresponding to ordinary inference rules such as *modus ponens*. Skeptical worries about rule-following in general will therefore be set aside, in order to focus on what is crucial about the metalanguage *ML*: namely that it includes *infinitary* inference rules. In the following sections, I will argue that it is *not* plausible to suppose that we have dispositions corresponding to such nonrecursive rules, and show that this turns Beth's Carnap\* into a real challenge.

#### 4.b Against infinitary rules

The argument to come can be summed up as follows: even if we grant that linguistic dispositions can be coordinated with inference rules like *modus ponens*, this strategy does not suffice to substantiate the claim that we speak *ML* rather than *ML\**. This is because these metalanguages involve nonrecursive inference rules, and, given plausible and widely shared assumptions, there are no linguistic dispositions that such rules could correspond to. Consequently only languages with recursive rules can be used *as metalanguages*—a restriction which undermines Carnap's approach to the philosophy of mathematics.

Let us revisit Carnap's claim that "there is nothing to prevent the practical application of" infinitary rules such as the  $\omega$ -rule, which we encountered in section 2.b. Unfortunately, Carnap does not elaborate on what exactly he means by *practical application* in this context, but if we take it to mean that we can *use* a metalanguage with infinitary rules just as easily as one with recursive rules, then his claim seems false. Practical application requires that we are, or at least could be, disposed to

<sup>9</sup>Carnap's most important papers on disposition are those from 1936, 1937b, and 1956b; for a survey of his position see Schmitz (2007). Williamson (2013, sec. 2.4) is a helpful introduction to Carnap on modality.

infer in accordance with something like the  $\omega$ -rule, and, for finite creatures such as ourselves, this appears to be impossible.

For one thing, since we will never be in a position to encounter infinitely many premises, there can be no cases in which we *actually* follow the rule. This in itself casts severe doubt on the claim that we may be *disposed* to do so, since other inference rules such as *modus ponens* are actually used in practice. Furthermore, consider what has been called the *Cognitive Constraint*: “humans cannot be attributed noncomputational causal powers” (Warren and Waxman *forthcoming*, 485). It is motivated by Van McGee in the following way:

Human beings are products of nature. They are finite systems whose behavioral responses to environmental stimuli are produced by the mechanical operation of natural forces. Thus, according to Church’s Thesis, human behavior ought to be simulable by a Turing machine. This will hold even for idealized humans who never make mistakes and who are allowed unlimited time, patience, and memory. (1991, 117)

Given Carnap’s sympathies with behaviourism, it is very likely that he would have endorsed the Cognitive Constraint. But there is wide agreement in the literature that this constraint prevents human beings from using nonrecursive inference rules since no Turing machine can decide whether they have been correctly applied or not (Field 1994; Raatikainen 2005; Button and Walsh 2018, chap. 7). This consensus has only very recently been challenged by Warren, who argues that following the  $\omega$ -rule does not require nonrecursive abilities after all (2020a; 2020b). A full assessment of his account of infinite reasoning, including a study of whether it is compatible with Carnap’s own philosophical commitments, goes beyond the scope of this paper. I will argue below, however, that even if Warren’s defence of the  $\omega$ -rule succeeds, it would not suffice to defend Carnap against a generalised version of Beth’s objection.

Scholars of Carnap usually accept that there are no infinitary dispositions straightforwardly corresponding to infinitary rules (Lavers 2004, 313; Ebbs 2017, 31). Ricketts therefore draws the conclusion that the connection between linguistic dispositions and a language with infinitary rules is a loose one:

We see, then, that as regards transformation rules, the agreement between a calculus and speech habits in virtue of which a language can be taken to instantiate a calculus is rather loose. For a speaker’s habits to agree with a calculus, Carnap appears to require little more than that the speaker not be disposed to affirm any contravalid [= false] sentence nor to deny any valid [= true] sentence. (2003, 262)

In light of these considerations, we can now revisit Carnap’s earlier response to Beth. Carnap maintained that we don’t have to worry about Carnap\* since we actually speak *ML* rather than *ML\**. But only now can we really appreciate what the claim that we speak *ML* amounts to: Carnap must hold that our speech dispositions correspond to *ML* rather than *ML\**. Is this the case? At first sight, one might think that the answer is *yes*. For  $\text{Con}_{PA}$  is true in *ML* and false in *ML\**. With few exceptions, everyone who understands what it means either accepts  $\text{Con}_{PA}$  or is neutral about its truth value. So by Ricketts’s criterion, *ML\** is not compatible with our dispositions.

This argument does not go far enough to help Carnap, however, for a number of reasons. First, one might think that the mere fact that  $\text{Con}_{PA}$  is widely accepted is not enough since it should also be accepted for the *right reasons*. To see what this means, remember that Beth described Carnap\*’s acceptance of  $\neg\text{Con}_{PA}$  as “psychopathic,” because it appears to be a brute conviction not supported by anything else. Presumably we want to say that the acceptance of  $\text{Con}_{PA}$  is different in character: not just a competing brute conviction that is more popular, but also one that is better *justified*. But it is not clear whether Carnap can say this. Arguably the disposition to accept  $\text{Con}_{PA}$  would somehow have to flow from the dispositions that constitute the acceptance of PA itself, but it is hard to see how this could be the case given the independence of  $\text{Con}_{PA}$  from the axioms of PA.

I think this is an uncomfortable conclusion, since if the acceptance of  $\text{Con}_{PA}$  is just as much a brute fact as the acceptance of  $\neg\text{Con}_{PA}$ , we come dangerously close to a view according to which the consistency of PA is a matter of choice, akin to the radical conventionalism that has been ascribed to Wittgenstein (Dummett 1959). I am happy to admit that this worry is not a knock-down argument, however, and that the response considered seems coherent in itself. But it is important to stress that Ricketts's proposal only provides a defence of Carnap if we construe Beth's objection in a narrow way, namely as concerning the truth value of  $\text{Con}_{PA}$  specifically. It is plausible to interpret Carnap's claim that mathematics is analytic in a broader sense, however, which is why we will turn to two kinds of generalisations next.

#### 4.c Beth's argument generalised

The sentence  $\text{Con}_{PA}$  we have been focussing on is just one example of an undecidable sentence, even though an especially interesting one. In addition, there are *infinitely many* other purely mathematical sentences which are independent of the axioms of PA, and for most of them we will have no inclination to either assert or deny them. So even if our speech behaviour does exclude the specific metalanguage  $ML^*$  Beth describes, there will still be infinitely many alternative metalanguages  $ML^{**}$ ,  $ML^{***}$ , etc., that are compatible with it, and which differ concerning the truth value of some undecidable sentence when compared to the standard model. Using Ricketts's criterion, one can thus at most argue that we do *not* speak  $ML^*$ , but we cannot draw the further conclusion that we *do* speak  $ML$  rather than one of the infinitely many other nonstandard metalanguages.

Against this, one might object that there is no need to exclude all of the deviant metalanguages since, while we care about the truth value of  $\text{Con}_{PA}$ , the vast majority of undecidable sentences are of no particular interest and so one could live with indeterminate truth values in such cases. But I do not think this is a satisfactory reply for both exegetical and systematic reasons. On the exegetical side, Carnap nowhere suggests that he only considers *some* undecidable sentences to be analytic (or contradictory). He usually describes mathematics as consisting of analytic and contradictory sentences exclusively, without a third category of sentences that are indeterminate or true without being analytic (1937a, 116). And this does not seem to be an incidental component of Carnap's position either because a mathematical sentence that is neither analytic nor contradictory would have to be classified as *synthetic*, a category that Carnap reserves for claims about the empirical world (1937a, 41). If he wants to retain these commitments, then the rules of the metalanguage need not only settle the truth value of one but of *all* undecidable sentences, and so all of the infinitely many deviant metalanguages  $ML^{**}$ ,  $ML^{***}$ , etc., need to be excluded.

On the systematic side, consider *why* we think that the truth value of  $\text{Con}_{PA}$  is important. Presumably the reason is that it expresses a claim about *syntax*, and there is a strong presumption to think that it must be determinate whether a contradiction is derivable from some axioms given some rules or not. But this seems to hold not just for the special case of PA but for formal theories in general, which makes the task at hand much harder. For the set of true consistency sentences, despite being a proper subset of all true mathematical claims, is itself nonrecursive (Clarke-Doane 2020, 161). The need for nonrecursive dispositions thus reappears even if all we require of Carnap is to explain the determinacy of claims about syntactic consistency, which seems a moderate and reasonable demand.

Another kind of generality problem emerges once we look beyond the case of Peano arithmetic. It is plausible enough that we accept  $\text{Con}_{PA}$  since the consistency of PA is regarded as well-established. But PA was just a convenient example, and it is clear that Carnap also wants to apply his approach to other mathematical theories, including set theory, and also theories concerning other abstract objects such as propositions (1956a, 205). In order to treat Zermelo-Fraenkel set theory (ZFC) in analogy to the case of PA, for instance, he needs to claim that we use a metalanguage which settles the consistency sentence  $\text{Con}_{ZFC}$ , and in general for any theory  $T$  we want to use, we would need to speak a metalanguage in which  $\text{Con}_T$  is true. But for many logical systems, such as ZFC plus

some large cardinal axiom, it will be an open question whether the system is consistent or not. Those who are well-informed and epistemically responsible about such matters will presumably have no disposition to either assert or deny the relevant consistency sentence.

This observation also demonstrates the limits of Warren's defence of the  $\omega$ -rule which I mentioned above. Adding the  $\omega$ -rule to PA does result in a complete theory, and hence there is at least some hope to save the analyticity of *arithmetic* from Beth's objection. But there is no analogous rule to the  $\omega$ -rule for ZFC, or most other formal theories, and so even in the best case the unified treatment of mathematical discourse Carnap aimed for is untenable. We can thus conclude that Carnap's position faces serious obstacles *even if* we make very generous assumptions about what our linguistic dispositions can commit us to, many of which may themselves be questioned. This shows that Beth's argument is much more forceful than has been recognised so far.

#### 4.d Incompleteness revisited

The dispute between Carnap and Beth revolves around the use of metalanguages, and it is useful to distinguish two distinct roles Carnap has for them. In the first instance, they are needed to make communication possible:

##### MEANS OF COMMUNICATION

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

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

##### RESOLVES INCOMPLETENESS

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

This move was supposed to capture the idea that mathematical truth is a matter of linguistic rules. Carnap seems to have thought that using strong metalanguages for this task is unproblematic, since the only objections that could arise are of a metaphysical kind, and hence in tension with the principle of tolerance. But Beth's argument shows that this is too quick, for when a formal language is actually to be *used* as a metalanguage, there needs to be a sense in which our linguistic dispositions can correspond to the rules of this language. And this requirement severely limits what languages can be employed in this role.

It seems that Carnap faces a dilemma: resolving incompleteness requires strong metalanguages, but communication requires weak metalanguages. Is there a way out? One idea would be to have distinct metalanguages play the different roles. Maybe what we actually speak is a weak metalanguage, but a strong metalanguage such as *ML* can still resolve incompleteness. But this raises the question of what *privileges ML* over deviant metalanguages if not the fact that *ML* is the one we use. For consider the lesson Beth himself, but also Stephen Kleene, drew from the case of undecidable sentences: namely that since some metalanguages make them true and others make them false, linguistic rules do *not* determine their truth values in the sense of, say, " $2 + 2 = 4$ " (Kleene 1939, 84f; Beth 1963, 478). This conclusion seems unavoidable unless one can argue that *ML*, and not some alternative metalanguage, encodes the correct rules when it comes to mathematics. But Carnap does not seem to have any resources left to do so: by the principle of tolerance he cannot appeal to metaphysical considerations to establish the superiority of *ML*, and by Beth's argument he cannot appeal to the fact that we actually speak *ML* either. It is hard to see what other option there is.

Alternatively, Carnap could try to challenge the inference from “our dispositions don’t pin down *ML* as our metalanguage” to “we do not speak *ML*”. Maybe the way I presented the coordination between natural and formal languages was too simple, or the assumption that dispositions can only correspond to recursive inference rules is incorrect. I am happy to admit that there *may* be room to manoeuvre here, but remain skeptical until a concrete proposal is on the table. What I take to be clear, however, is that Carnap’s actual response—namely to insist that we speak *ML* as if this were completely unproblematic—is not adequate, despite being frequently endorsed. Defenders of Carnap need to say more.

## 5. Beth and model-theoretic arguments

### 5.a Varieties of nonstandard models

This last section will wrap up the paper by coming back to an exegetical question I set aside earlier: namely whether the argument I presented is really *Beth’s* argument from nonstandard models rather than *my own*. This is especially salient since Beth never explicitly mentions dispositions, which played such a crucial role in my reconstruction. I will give some reasons for thinking that Beth must nevertheless have had a version of the argument against nonrecursive inference rules in mind.

It will be illuminating to approach this issue by comparing Beth’s objection to some other *model-theoretic* arguments that are well-known. Let us first draw some explicit distinctions between different kinds of nonstandard models. Taking  $\mathbb{N}$  to be the standard model of PA, there are nonstandard models whose domain is isomorphic to  $\mathbb{N}$  but contains different objects:

#### ISOMORPHIC

A model which is isomorphic but not identical to the standard model.

*Example:* Interpreting PA in the sets of even numbers by mapping each number  $n$  to  $2n$ .

If we just care about the truth and falsity of the sentences in the language of PA, it does not matter whether we interpret the theory in  $\mathbb{N}$  or an isomorphic nonstandard model: the various models will make the very same sentences true and false, a property which is called *elementary equivalence*. It is a consequence of the Löwenheim-Skolem theorem that there are also models which are elementarily equivalent to  $\mathbb{N}$  but *not* isomorphic to it. This brings us to the second kind of nonstandard model:

#### NONISOMORPHIC BUT ELEMENTARILY EQUIVALENT

A model which is not isomorphic to the standard model but nevertheless makes the very same sentences true and false.

*Example:* Interpreting PA in an uncountable model via the upward Löwenheim-Skolem theorem.

The third kind of nonstandard model is what Beth’s argument crucially relies on, namely one in which a sentence of PA has a different truth value compared to the standard model:

#### TRUTH-SWITCHING

A model which is not isomorphic to the standard model and changes the truth value of some sentence from those of the standard model.

*Example:* Interpreting PA in a model where  $\neg \text{Con}_{PA}$  holds.

Not all philosophical uses of model-theoretic considerations rely on truth-switching nonstandard models. At least one version of Hilary Putnam’s famous model-theoretic argument against metaphysical realism, for instance, only requires the existence of isomorphic nonstandard models. Putnam attacks a certain view about the relationship between language and the world, according to which “THE WORLD” is in itself divided into discrete objects, which are in turn linked to the words of our language by a reference relation (1977, 483f).

The underlying picture, according to which there is some kind of one-to-one correspondence between words and entities, is a natural one, but Putnam argues that it faces a serious problem. Thinking about the situation model-theoretically, the metaphysical realist wants to say that THE WORLD is in effect the intended interpretation of our best theory of the world. But as Putnam points out, any theory which has a model at all has *many*:

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

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

I will not pursue this issue further here, as there is little temptation to interpret Carnap as a metaphysical realist.<sup>10</sup> But there is another kind of model-theoretic argument which also does not require truth-switching nonstandard models but has in fact been ascribed to Beth, and we will turn to it next.

### 5.b Skolem's paradox

Right after describing Carnap\*, Beth writes that “the above considerations [...] are only variants of the Löwenheim-Skolem paradox” (1963, 478). Ricketts has taken this reference very seriously, and interprets Beth's argument as being a version of what is usually called *Skolem's paradox* (2004). In a nutshell, the (alleged) paradox goes as follows: in ZFC, it is possible to prove that there are uncountable sets. But, by the downward Löwenheim-Skolem theorem, we know that ZFC has countable models. And one might worry that this result somehow undermines the claim that the sentence “there are uncountable sets” expressed within ZFC actually means that there are uncountable sets (Tymoczko 1989). Analogously, so the suggestion, one could read Beth as worrying that the possibility of nonstandard models somehow shows that we can't really talk about the natural numbers using PA.

Since, despite its name, Skolem's paradox is not usually regarded as a genuine paradox, this analogy would make Beth's argument easy to defuse.<sup>11</sup> The correct response to this, however, is not to dismiss Beth's argument as ineffective, but rather to set aside his comparison to Skolem's paradox as misleading. For consider how, in his reconstruction, Ricketts describes the case of Carnap and Carnap\*:

One logician might understand the arithmetic in the informal syntax language standardly; the other might understand it non-standardly. *This divergence need not be manifest in their use of the sentences of the informal syntax language.* [...] Having read our Skolem, we observe that we can state transformation rules in two different ways, one corresponding to the standard model of arithmetic, another corresponding to a non-standard model. *We may suppose that each group of transformation rules demarcates the same formulas of the object calculus as true.* (2004, 194; my emphasis)

This scenario is certainly in keeping with the Löwenheim-Skolem theorem, for it only proves the possibility of what I called nonisomorphic but elementarily equivalent nonstandard models. But,

<sup>10</sup>See Button (2013) for a comprehensive discussion of Putnam's model-theoretic argument.

<sup>11</sup>See Bays (2014) for an overview of the most common philosophical responses.

much more importantly, it does not match Beth's own description of the case, in which the difference between Carnap and Carnap\* is manifested in their attitudes towards certain sentences, such as  $Con_{PA}$ . My interpretation, which relies on the possibility of truth-switching nonstandard models, thus makes better sense of the crucial part of Beth's paper than that of Ricketts.

While these considerations support the exegetical accuracy of my reading, they do not suffice to fully dispel the worry I raised earlier on: namely that, unlike me, Beth nowhere explicitly talks about dispositions and their correspondence to nonrecursive rules. It is possible to alleviate this concern in an indirect way, however, since an otherwise very puzzling passage makes sense once it is read in light of my interpretation. After discussing other aspects of Carnap's philosophy, Beth comes back to the topic of strong metalanguages toward the end of his paper:

In the first place, such a language, which may be called  $M$ , should enable us, as pointed out by Church, to state the necessary directives for the concrete manipulation of certain physical objects, namely, the signs of the object language. This implies that  $M$  must contain the means of expression for a certain version of elementary arithmetic or of a suitable general arithmetic. Moreover, this part of  $M$ , which will be called  $M_1$ , must be *understood in accordance with strict usage*. This demand, however, strongly restricts the development of  $M_1$ , the language of elementary syntax, into a means of expression, called  $M_2$ , for theoretical syntax; this follows from our discussion [of Carnap and Carnap\*]. (1963, 499f)

Here  $M_1$  must be some version of arithmetic in order to encode claims about syntax, and Beth seems to say that there is no problem with understanding  $M_1$  in accordance with strict usage. But prima facie this is quite mysterious, for the incompleteness theorems also apply to very weak theories of arithmetic such as Robinson arithmetic, and so they too will have truth-switching nonstandard models.<sup>12</sup> And Beth's definition I cited in section 3.a suggested that *strict usage* requires pinning down one particular model, which still seems to be a problem even for  $M_1$ .

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

The upshot of this section can be described as follows: Despite being called the argument from nonstandard models, there are both systematic and exegetical reasons to interpret Beth's objection in a way that makes it more similar to the underdetermination arguments of Quine and Kripke rather than the model-theoretic arguments of Putnam and Skolem. One further advantage of this reading is that worries about anachronism can be set aside. After all, in *Logical Syntax*, Carnap still rejected semantic notions such as *truth* and *reference*, and hence a model-theoretic conception of logic seems quite alien to this syntactic outlook.<sup>13</sup> This would be an additional challenge for interpretations of Beth along Ricketts's lines, but does not affect the problem of dispositions corresponding to nonrecursive rules that I stress.

<sup>12</sup>Beth will have known about this since he was one of the proofreaders of the influential *Undecidable Theories* by Tarski, Mostowski, and Robinson (1953, ix), in which this result is proofed.

<sup>13</sup>It has been convincingly argued, however, that what Carnap describes as syntax is effectively semantics in disguise, since his definition of analyticity for language II is a variant of a semantic truth-definition à la Tarski (Coffa 1987; Koellner 2009). Soon after *Logical Syntax*, Carnap openly embraced semantic notions, and later presentations of his view include model-theoretic ideas, such as the specification of an intended interpretation for an object language (1963, 900f). For a discussion of the continuities and differences between Carnap's syntactic and semantic period see Ricketts (1996) and Goldfarb (1997).

Presumably this will not convince everyone, and some might object that there is now *too little* model-theory in my reconstruction. But as it stands, I regard my reading of Beth as being in good shape because it is coherent, effective, and fits most of the original text.

## 6. Conclusion

I think that my investigation of Beth's argument illustrates a more general point about Carnap's philosophy. The description and development of formal languages, with a focus on their rules, takes up a lot of space in Carnap's books and papers. Remarks on what it means to adopt and use such languages and how they relate to natural language more generally, however, are very rare. This lack of interest may partly explain why Carnap did not engage more fully with Beth's argument, for if we restrict our attention to the formal systems as such, everything seems fine. It is then also no surprise that those of Carnap's contemporaries who rejected the sharp distinction between formal systems and natural language—namely Neurath<sup>14</sup> and Quine<sup>15</sup>—found themselves at odds with Carnap, but struggled to get their objections across. I hope that my discussion has shown how valuable it can be to ask the questions that Carnap himself preferred to set aside. While it would be nice if the principle of tolerance on its own could dissolve all the challenges Carnap's philosophy of mathematics faces, thanks to Beth we now know that the situation is not that simple.

**Acknowledgements.** This paper is based on my PhD thesis, and I thank Tim Button for all his advice and encouragement during the years of writing it. Another special thanks goes to Tom Ricketts for illuminating discussions about Carnap and Beth while I visited the University of Pittsburgh in early 2020. I furthermore thank the two referees for this journal, whose helpful and constructive comments led to numerous improvements, as well as everyone who read earlier drafts: Christian Feldbacher-Escamilla, Makan Nojournian, Paula Keller, Sander Verhaegh, Sophia Arbeiter, and Stephen Mackereth. My research was funded by the Arts and Humanities Research Council, the Cambridge Trust, the Aristotelian Society, the Royal Institute of Philosophy, and the German-American Fulbright Commission.

**Benjamin Marschall** recently completed a PhD on Carnap and the ontology of mathematics at the University of Cambridge. He is currently a visiting fellow at the Institute Vienna Circle working on Quine's philosophy of mathematics.

## References

- Awodey, Steve, and André W. Carus. 2004. "How Carnap Could Have Replied to Gödel." In *Carnap Brought Home: The View from Jena*, edited by Steve Awodey and Carsten Klein, 203–23. Chicago: Open Court.
- Bays, Timothy. 2014. "Skolem's Paradox." In *The Stanford Encyclopedia of Philosophy* (Winter 2014), edited by Edward N. Zalta. <https://plato.stanford.edu/entries/paradox-skolem/>.
- Beth, Evert Willem. 1963. "Carnap's Views on the Advantages of Constructed Systems over Natural Languages in the Philosophy of Science." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 469–502. La Salle, IL: Open Court.
- Buldt, Bernd. 2004. "On RC 102-43-14." In *Carnap Brought Home: The View from Jena*, edited by Steve Awodey and Carsten Klein, 225–46. Chicago: Open Court.
- Button, Tim. 2013. *The Limits of Realism*. Oxford: Oxford University Press.
- Button, Tim, and Sean Walsh. 2018. *Philosophy and Model Theory*. Oxford: Oxford University Press.
- Carnap, Rudolf. 1936. "Testability and Meaning." *Philosophy of Science* (34): 419–71.
- Carnap, Rudolf. 1937a. *The Logical Syntax of Language*. London: K. Paul, Trench, Trubner & Co.
- Carnap, Rudolf. 1937b. "Testability and Meaning—Continued." *Philosophy of Science* (41): 1–40.

<sup>14</sup>According to Thomas Uebel (1992), the disagreement between Carnap and Neurath on private languages results from the fact that Neurath—unlike Carnap—regarded it as crucial to construe languages in such a way that human beings can actually speak them.

<sup>15</sup>Quine (1963) famously rejected the central doctrine of Carnap's philosophy of mathematics: namely that the possibility of defining mathematical truth in a strong metalanguage shows it to be a matter of linguistic rules. Although the details are complex, part of Quine's reason is that he regards formal and natural languages as continuous and thinks that we cannot give a definition of "analytic" for natural languages (Quine and Carnap 1990, 438f; Ricketts 2011, 288f; Verhaegh 2018, 74f).

- Carnap, Rudolf. 1939. "Foundations of Logic and Mathematics." In *International Encyclopedia of Unified Science*, edited by Otto Neurath, Rudolf Carnap, and Charles Morris, C, 139–213. Chicago: University of Chicago Press.
- Carnap, Rudolf. 1955. "Meaning and Synonymy in Natural Languages." *Philosophical Studies* 6 (3): 33–47.
- Carnap, Rudolf. 1956a. "Empiricism, Semantics, and Ontology." In *Meaning and Necessity*, 205–221. Chicago: University of Chicago Press.
- Carnap, Rudolf. 1956b. "The Methodological Character of Theoretical Concepts." *Minnesota Studies in the Philosophy of Science* 1 (1): 38–76.
- Carnap, Rudolf. 1963. "Replies and Systematic Expositions." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 859–1013. La Salle, IL: Open Court.
- Carnap, Rudolf. 1983. "The Logicist Foundations of Mathematics." In *Philosophy of Mathematics: Selected Readings*, edited by Paul Benacerraf and Hilary Putnam, 41–52. Cambridge: Cambridge University Press.
- Carus, André W. 2004. "Sellars, Carnap, and the Logical Space of Reasons." In *Carnap Brought Home: The View from Jena*, edited by Steve Awodey and Carsten Klein, 317–55. Chicago: Open Court.
- Carus, André W. 2007. *Carnap and Twentieth-Century Thought: Explication as Enlightenment*. Cambridge: Cambridge University Press.
- Clarke-Doane, Justin. 2020. *Morality and Mathematics*. New York: Oxford University Press.
- Coffa, J. Alberto. 1987. "Carnap, Tarski and the Search for Truth." *Noûs* 21 (4): 547–72.
- Dummett, Michael. 1959. "Wittgenstein's Philosophy of Mathematics." *Philosophical Review* 68 (3): 324–48.
- Ebbs, Gary. 2011. "Carnap and Quine on Truth by Convention." *Mind* 120 (478): 193–237.
- Ebbs, Gary. 2017. "Carnap's Logical Syntax." In *Carnap, Quine, and Putnam on Methods of Inquiry*, 13–32. New York: Cambridge University Press.
- Field, Hartry. 1994. "Are Our Logical and Mathematical Concepts Highly Indeterminate?" *Midwest Studies in Philosophy* 19 (1): 391–429.
- Flocke, Vera. 2019. "Carnap's Defense of Impredicative Definitions." *The Review of Symbolic Logic* 12 (2): 372–404.
- Friedman, Michael. 1999a. "Analytic Truth in Carnap's *Logical Syntax of Language*." In *Reconsidering Logical Positivism*. New York: Cambridge University Press.
- Friedman, Michael. 1999b. "Tolerance and Analyticity in Carnap's Philosophy of Mathematics." In *Reconsidering Logical Positivism*, 198–233. New York: Cambridge University Press.
- Friedman, Michael. 2009. "Tolerance, Intuition, and Empiricism." In: *Carnap's Logical Syntax of Language*, edited by Pierre Wagner, 236–49. New York: Palgrave-Macmillan.
- Gödel, Kurt. 1995. "Is Mathematics Syntax of Language?" In *Kurt Gödel Collected Works. Volume III. Unpublished Essays and Lectures*, edited by Solomon Feferman, John W. Dawson Jr., Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 334–55. Oxford: Oxford University Press.
- Goldfarb, Warren. 1995. "Introductory Note to \*1953/9." In *Kurt Gödel Collected Works. Volume III. Unpublished Essays and Lectures*, edited by Solomon Feferman, John W. Dawson Jr., Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 324–33. Oxford: Oxford University Press.
- Goldfarb, Warren. 1997. "Semantics in Carnap: A Rejoinder to Alberto Coffa." *Philosophical Topics* 25 (2): 51–66.
- Goldfarb, Warren, and Thomas Ricketts. 1992. "Carnap and the Philosophy of Mathematics." In *Wissenschaft und Subjektivität. Science and Subjectivity*, edited by David Bell and Wilhelm Vossenkuhl, 61–78. Berlin: Akademie Verlag.
- Kleene, Stephen C. 1939. "Review: Rudolf Carnap, *The Logical Syntax of Language*." *Journal of Symbolic Logic* 4 (2): 82–87.
- Koellner, Peter. 2009. "Carnap on the Foundations of Logic and Mathematics." Unpublished manuscript. <http://logic.harvard.edu/koellner/CFLM.pdf>.
- Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
- Lavers, Gregory. 2004. "Carnap, Semantics and Ontology." *Erkenntnis* 60 (3): 295–316.
- Lavers, Gregory. 2019. "Hitting a Moving Target: Gödel, Carnap, and Mathematics as Logical Syntax." *Philosophia Mathematica* 27 (2): 219–43.
- McGee, Van. 1991. "We Turing Machines Aren't Expected-Utility Maximizers (Even Ideally)." *Philosophical Studies* 64 (1): 115–23.
- Potter, Michael. 2000. *Reason's Nearest Kin: Philosophies of Arithmetic from Kant to Carnap*. Oxford: Oxford University Press.
- Putnam, Hilary. 1977. "Realism and Reason." *Proceedings and Addresses of the American Philosophical Association* 50 (6): 483–98.
- Putnam, Hilary. 1980. "Models and Reality." *Journal of Symbolic Logic* 45 (3): 464–82.
- Putnam, Hilary. 1983. "Philosophers and Human Understanding." In *Realism and Reason*, 184–204. Cambridge: Cambridge University Press.
- Putnam, Hilary. 1994. "Thomas Ricketts on Carnap." In *Reading Putnam*, edited by Peter Clark and Bob Hale, 280–81. Oxford: Blackwell.
- Quine, Willard Van. 1949. "Truth by Convention." In *Reading in Philosophical Analysis*, edited by Herbert Feigl and Wilfrid Sellars, 250–73. New York: Appleton-Century-Crofts.
- Quine, Willard Van. 1960. *Word and Object*. Cambridge, MA: MIT Press.

- Quine, Willard Van. 1963. "Carnap and Logical Truth." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 385–406. La Salle, IL: Open Court.
- Quine, Willard Van, and Rudolf Carnap. 1990. *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*, edited and with an introduction by Richard Creath. Berkeley: University of California Press.
- Raatikainen, Panu. 2005. "On Horwich's Way Out." *Analysis* 65 (3): 175–77.
- Raatikainen, Panu. 2020. "Gödel's Incompleteness Theorems." In *The Stanford Encyclopedia of Philosophy* (Summer 2020), edited by Edward N. Zalta. <https://plato.stanford.edu/entries/goedel-incompleteness/>.
- Richardson, Alan W. 1994. "Carnap's Principle of Tolerance." *Proceedings of the Aristotelian Society* 67: 67–82.
- Richardson, Alan W. 2012. "Carnap's Place in Analytic Philosophy and Philosophy of Science." In *Carnap's Ideal of Explication and Naturalism*, edited by Pierre Wagner. New York: Palgrave-Macmillan.
- Ricketts, Thomas. 1994. "Carnap's Principle of Tolerance, Empiricism, and Conventionalism." In *Reading Putnam*, edited by Peter Clark and Bob Hale, 176–200. Oxford: Blackwell.
- Ricketts, Thomas. 1996. "Carnap: From Logical Syntax to Semantics." In *Origins of Logical Empiricism*, edited by Ronald N. Giere and Alan W. Richardson, 231–50. Minnesota: University of Minneapolis Press.
- Ricketts, Thomas. 2003. "Languages and Calculi." In *Logical Empiricism in North America*, edited by Gary L. Hardcastle and Alan W. Richardson, 257–80. Minneapolis: University of Minnesota Press.
- Ricketts, Thomas. 2004. "Frege, Carnap, and Quine: Continuities and Discontinuities." In *Carnap Brought Home: The View from Jena*, edited by Steve Awodey and Carsten Klein, 181–202. Chicago: Open Court.
- Ricketts, Thomas. 2007. "Tolerance and Logicism: Logical Syntax and the Philosophy of Mathematics." In *The Cambridge Companion to Carnap*, edited by Michael Friedman and Richard Creath, 200–25. Cambridge: Cambridge University Press.
- Ricketts, Thomas. 2011. "Roots of Ontological Relativity." *American Philosophical Quarterly* 48 (3): 287–300.
- Schmitz, François. 2007. "Dispositions and Counterfactuals. From Carnap to Goodman's Children and Grandchildren." In *Dispositions and Causal Powers*, edited by Max Kistler and Bruno Gnassounou, 43–66. London: Routledge.
- Tarski, Alfred, Andrzej Mostowski, and Raphael M. Robinson. 1953. *Undecidable Theories*. Amsterdam: North-Holland Publishing Company.
- Tymoczko, Thomas. 1989. "In Defense of Putnam's Brains." *Philosophical Studies* 57 (3): 281–97.
- Uebel, Thomas. 1992. "Neurath vs. Carnap: Naturalism vs. Rational Reconstructionism before Quine." *History of Philosophy Quarterly* 9 (4): 445–70.
- Verhaegh, Sander. 2018. *Working from Within: The Nature and Development of Quine's Naturalism*. New York: Oxford University Press.
- Wagner, Pierre. 2012. "Natural Languages, Formal Systems, and Explication." In *Carnap's Ideal of Explication and Naturalism*, edited by Pierre Wagner, 175–89. New York: Palgrave-Macmillan.
- Warren, Jarred. 2015. "Conventionalism, Consistency, and Consistency Sentences." *Synthese* 192 (5): 1351–71.
- Warren, Jarred. 2020a. *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. New York: Oxford University Press.
- Warren, Jarred. 2020b. "Infinite Reasoning." *Philosophy and Phenomenological Research*. <https://doi.org/10.1111/phpr.12694>.
- Warren, Jarred, and Daniel Waxman. forthcoming. "A Metasemantic Challenge for Mathematical Determinacy." *Synthese* 197 (2): 477–95.
- Williamson, Timothy. 2013. *Modal Logic as Metaphysics*. Oxford: Oxford University Press.