

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

Casper Storm Hansen. *Founding Mathematics on Semantic Conventions*. Synthese Library; 442. Springer, 2021.

Reviewed by **Brett Topey***

To appear in *Philosophia Mathematica*. Please cite published version.

For two books that describe themselves as developing and defending a kind of conventionalism about logic and mathematics—and that seem to be motivated by broadly the same metaphysical and epistemological scruples—Casper Storm Hansen’s *Founding Mathematics on Semantic Conventions* and Jared Warren’s *Shadows of Syntax* could hardly be more different. Hansen’s project is a revisionary one: for him, classical mathematics commits us to the existence of mathematical objects we don’t have good reason to believe in and so must be given up. So his goal is to rebuild as much of classical mathematics as possible nominalistically, in a constructed language in which the mathematical terms aren’t referring terms at all, with truth conditions for mathematical sentences instead laid down via explicit stipulations: the ‘semantic conventions’ of the book’s title. Warren, on the other hand, relies on a radically permissive sort of inferentialist metasemantics, and its accompanying deflationism about ontology, to develop a picture according to which logic and mathematics as actually practiced can’t *help* but be correct. As will become clear, Warren’s project is, to my mind, the more successful. I’ll discuss the two books in turn.

Hansen begins in Chapter 1 by explaining what sorts of items he thinks can legitimately be appealed to in an account of mathematics. These are restricted to ‘entities and facts...we have reason to believe in, independently of mathematics’ (p. 2), and include physical objects and facts about them; thinking beings, thoughts, and facts about them (whether these are ultimately physical or not); and, most controversially, nontrivial facts about necessity and possibility. They don’t include (mind-independent) mathematical objects or any other (mind-independent) abstract objects, or even *possible* such objects, nor do they include possible sets that are genuinely (i.e., actually, in Aristotle’s sense) infinite—Hansen suggests that, while he’s ‘comfortable with basing mathematics on the possibility of...thinking beings having other thoughts than the ones they actually have, and thoughts that complexity-wise go far beyond what any actual human being is capable of’ (p. 4), the possibility

*Department of Humanities and Life Sciences, Scuola Universitaria Superiore IUSS Pavia, 27100 Pavia, Italy. [Orcid.org/0000-0002-4603-0569](https://orcid.org/0000-0002-4603-0569). Email: brett.topey@iusspavia.it.

of abstract objects and actual infinity is too doubtful to be relied on.¹ Hansen, then, sets out to devise as powerful a mathematical system as possible within these constraints.

In Chapters 2–4 he surveys some existing approaches to mathematics, finding that each of them either relies on illegitimate ontological assumptions or else fails to fully exploit the resources that are legitimately available and so is less useful than it might be. Chapter 2 contains his critique of the sort of relaxed attitude toward set existence—a ‘plenitudinous combinatorialist’ attitude—that has characterized mainstream mathematics since Cantor. In particular, Hansen tries to show that set-theoretic arguments fail to give us reason to accept the (actual or possible) existence of certain combinatorial entities—undefinable sets—commitment to which would, by Hansen’s lights, involve commitment to abstract objects and actual infinity.² (More on this critique below.) Chapter 3 targets Brouwer’s intuitionism; here Hansen criticizes Brouwer’s strategy of characterizing undefinable real numbers by appeal to possible infinitely long choice sequences that are ‘freely proceeding’ (i.e., aren’t governed by any algorithm), on the grounds that, *contra* Brouwer, this strategy requires actual infinity in order to be successful, since no choice sequence can pick out a particular undefinable real number until that sequence is complete. And Chapter 4 discusses Frege’s logicism and some of its descendants. These approaches fare better than the previous ones, according to Hansen: their notion of set is one on which sets are definable by their very nature, since a set is just the extension of a predicate. But Frege’s own system is of course inconsistent, and the fix made available by Russell’s ramified type hierarchy would, Hansen suggests, require ‘platonic predicates beyond the grasp of humans’ (p. 57), since Russell’s axiom of reducibility, which makes his reconstruction of classical mathematics possible, ‘would likely be false’ if the only predicates available for use were ‘mundane predicates that can be

¹Hansen never fully explains what’s behind his doubts about actually infinite sets, but the worry appears to be something like this: although *potentially* infinite sequences can (setting aside worries about rule-following) be accounted for by appeal to thought, it’s not possible to complete the process, in thought, of ‘gathering up’ infinitely many objects into a set, in which case actually infinite sets would, by Hansen’s lights, have to be mind-independent. If this is right, then despite what the structure of Hansen’s discussion suggests, his worry about actual infinity isn’t ultimately distinct from his worry about abstract objects.

²The reason this commitment to abstract objects is supposed to follow is that accepting the existence of a set that ‘defies definition and therefore cannot be identified with a thought’ requires assuming that if some objects exist, then (in nonparadoxical cases) the set of them also exists, ‘automatically and in the absence of any action or creation process’, and so, by Hansen’s lights, exists mind-independently (p. 2). (As for why the commitment to actual infinity is supposed to follow: if a set is undefinable, there’s no rule that can generate a list of its members, which means it can’t be accounted for just by appeal to potential infinity. Compare the below point about Brouwer’s strategy for characterizing undefinable real numbers.) It seems to me that it’s a mistake to identify existence in the absence of a creation process with mind-independence of any genuine sort, for reasons that will emerge below.

formed in a human language' (p. 56)—here the thought is that, if we are restricted to ordinary predicates, there plausibly will be real numbers that we can define at higher orders of type 1 but not at the first order. And although Weyl's more restrictive form of predicativism avoids this problem (and indeed is legitimate by Hansen's lights), it does so by eliminating higher orders altogether, in which case there plausibly are certain 'bona fide real numbers' that Weyl's system 'arbitrarily exclude[s]' (p. 65). Hansen suggests that a system based not on a *fixed* language but on an *open-ended* one—i.e., one such that 'new terms and forms of sentences can be added to it over time' (p. 62)—can (potentially) include the real numbers in question without reintroducing a Russellian hierarchy. He spends the rest of the book explaining and developing a system of this sort.

Chapter 5 is devoted to the liar paradox and the conventionality of logic. There's more going on in this chapter than can be dealt with here, but the main idea that's relevant for the overall project is that nonclassical semantic theories that have been deployed in response to the liar paradox should be understood simply as alternative language conventions we might adopt so as to handle defective sentences. Chapter 6 begins to clarify Hansen's nominalistic strategy for reconstructing mathematics via conventional stipulation of truth conditions for mathematical sentences, where some of these sentences are simply stipulated to be 'true outright' (p. 95)—i.e., true in *all* conditions—and, along the way, lays out Hansen's responses to Quine's regress argument against conventionalism and to Kripkensteinian worries about rule-following. (His response to worries about rule-following is simply to *assume* that humans are able to be committed to particular rules, while granting that 'there seems to be no way to reduce this ability...to simpler and more well-understood abilities or aspects of the world' (p. 107). See below for discussion of his response to Quine's regress argument.) Chapters 7–9 are largely technical; in these chapters Hansen introduces the Kripke-inspired trivalent semantics that serves as the foundation for his system (where constructing this semantics involves quantifying over possible languages—this is one reason he needs to insist on the legitimacy of appeals to nontrivial modal facts), lays out semantic stipulations governing various bits of mathematical vocabulary, proves that the resulting system reconstructs classical theories of natural numbers, integers, and rational numbers, and shows that versions of various important theorems of classical real analysis can be proved in the system as well. Finally, Chapter 10 attempts to clarify some of the modal claims that Hansen has relied on throughout the book.

It may or may not be clear from the broad summary just given, but this is an idiosyncratic book. Consider, for example, the discussion in Chapter 6 of Quine's regress argument. Here things begin as we might expect: Hansen rehearses Quine's (1936) reasoning and acknowledges the correctness of his point about the impossibility of generating infinitely many truths via finitely many explicit stipulations. But then he concludes from this, without further ado, that there must be, in addition

to our conventions, some fixed set of ‘primitive rules’ that are ‘impossible...to successfully explain’ to anyone who doesn’t ‘already grasp them (at least implicitly)’ but that we are nevertheless (innately?) able to follow, where these rules are ‘not conventional’ but instead constitute the basic patterns of thought that are available to us as humans (pp. 101–2).

This is a strange lesson to take from Quine’s argument. For one thing, it’s profoundly un-Quinean—Quine himself was a skeptic about implicit rule-following and so certainly wouldn’t have been willing to posit rules that are primitive in Hansen’s sense. For another: insofar as we, unlike Quine, are *not* skeptics about implicit rule-following, the obvious conclusion to draw, if we wish to tell a conventionalist story, is simply that some *conventional* rules are implicit, in which case (as Quine himself acknowledges) there’s no regress problem after all. (This, incidentally, is the starting point of Warren’s response to Quine, which I discuss below.) But Hansen ignores this possibility completely, for reasons that are obscure to me. And even setting all of that aside, the fact is that Quine’s argument doesn’t on its own tell us anything about how or why we come to follow the rules we do; what it tells us is that a particular account of how logical sentences come to have their *truth values* fails.³

It’s also notable that Hansen, after positing these rules in response to Quine’s argument, goes on to criticize Quine’s main target—i.e., Carnap—for his ‘naivete’ in failing to incorporate such rules into his picture: though Carnap, Hansen says, got it right in claiming, for example, that neither classical nor intuitionist logic is uniquely correct, Carnap was nevertheless ‘naive not to include any discussion of the highly non-trivial reasons *why* [that claim] is true, namely that humans have possession of the primitive rules needed for a classical semantics in addition to those needed for an intuitionist semantics’ (p. 103fn). This criticism is bizarre for at least two reasons. First, there *is* a nontrivial Carnapian explanation of why the claim in question is true: there’s no sense in which the rules of a language can be correct or incorrect because there’s no external standard of correctness for them to meet—the rules themselves determine what correctness consists in. This is the revolutionary idea that underlies the principle of tolerance, which was the key element of Carnap’s whole philosophical approach; to appeal to this idea rather than

³I suspect that what’s going on is that Hansen (without explicitly saying so) is taking Quine’s argument to be a way of developing the so-called ‘adoption problem’, according to which it’s impossible to adopt logical rules *explicitly* (Kripke 2024, Padro (now Birman) 2015), and that, because of his views about rule-following, he takes *implicit* commitment to particular rules to be primitive. But (setting aside the fact that the adoption problem is a problem Kripke develops for Quine’s own view of logic and so is certainly not the problem Quine took himself to be developing) even on the assumption that humans’ ability to follow rules is indeed primitive, it’s not clear why it’s supposed to follow that what’s primitive is only the ability to follow a particular *fixed* set of rules. In any case, I’m just guessing about Hansen’s picture here; none of these connections is made explicit in the text.

to the primitive rules posited by Hansen is hardly naive, especially since, again, positing such rules is not the natural response to Quine's argument. And second, even if we suppose Hansen is right that humans are in possession of some fixed set of primitive rules for thinking, this doesn't mean the rules for both classical and intuitionistic negation need to be among those rules in order for Carnap's claim to be true. Carnap (like Hansen) is primarily concerned with constructed languages. So, insofar as the metalanguage we're using to construct those languages—i.e., our natural language—has the expressive resources to describe both classical and intuitionistic rules, we can easily construct languages that include rules of either sort; whether both sorts of rules are among our primitive rules is neither here nor there. If, for instance, the primitive rules governing our natural language are classical, as Hansen suggests they are, it's going to be simple enough, in that language, to specify (syntactic or semantic) rules of a constructed language whose logic is intuitionistic.⁴

What's most striking about all this, though, is that outside of the footnote in which this criticism of Carnap is put forward, Hansen's book contains no mention at all of Carnap or his program. Given the amount of space Hansen devotes to discussing the inadequacy of other late-19th- and early-20th-century approaches to the philosophy of mathematics, this omission is odd; Carnap's motivations, after all, arguably are closer to Hansen's than are the motivations behind any of the approaches he *does* discuss. Indeed, it seems to me that Hansen's and Carnap's motivations are so similar that, if anything even in the neighborhood of Carnap's approach is workable, then the restrictions Hansen places on himself are entirely unnecessary. Given this fact, it's especially surprising that Hansen doesn't devote more attention to explaining what's wrong with a broadly Carnapian approach.

Which brings me to the main concern I want to discuss here. There's real formal ingenuity in Hansen's book—what he's able to do, given the limited resources he allows himself, is impressive. But in my view, he doesn't succeed in showing that the project of constructing a mathematics within such strict limitations is philosophically well motivated in the first place. In fact, I have my doubts about whether the metaphysical and epistemological scruples from which he begins can really give us any principled reason even to abandon classical mathematics, let alone to restrict ourselves to precisely the resources he takes to be available.

Let's suppose that we share Hansen's scruples (as in fact I do, more or less) and examine in this light some of the reasoning that leads him to proceed as he does. In particular, let's take a closer look at his critique in Chapter 2 of plenitudinous combinatorialism, which he describes both as a view on which 'sets can exist independently of minds, independently of a creation event, and independently of there

⁴Furthermore, there's nothing special about intuitionistic logic here; every nonclassical logic that has been seriously studied has rules that can be specified in a classical metalanguage.

being a property or predicate to “bind” their elements together’ and as a view on which we should place as few restrictions as we can, while avoiding contradiction, on the claim that ‘whenever there are some objects, there is a set of those objects’ (p. 9).

The first thing to note here is that plenitudinous combinatorialism, as Hansen characterizes it, pairs a methodological thesis about how set theory should be practiced with a metaphysical thesis about the nature of sets. Prima facie, these two theses seem separable in principle—it’s not clear why one would need to accept metaphysical platonism about abstract objects (or even its possibility—I’ll tend to leave this implicit from now on) in order to endorse any particular approach to the practice of set theory. Hansen, though, proceeds as though it’s obvious that there’s no space between these two commitments. So he focuses, in the bulk of the chapter, not on evaluating philosophical arguments concerning the ontology and metaontology of set theory but instead on arguing for the claim that certain mathematical results *within* set theory can’t provide genuine reason to believe in undefinable entities.

This claim is very plausibly correct, if we follow Hansen in construing mathematicians’ belief in undefinable entities as involving a commitment to metaphysical platonism. But its correctness, on this construal, is neither surprising nor particularly interesting, nor does establishing that correctness require examining the mathematical results themselves—all that’s needed is to note that, insofar as we take the conclusion of a set-theoretic proof to tell us something about how things are in platonic heaven, this can only be because we *already* took the basic set-theoretic principles involved in that proof to tell us something about how things are in platonic heaven.⁵ Despite this, Hansen *does* examine the relevant mathematical results, and here things again begin to look strange.

Here’s one argument he discusses for the existence of undefinable entities. By Cantor’s Theorem, the power set of the set of natural numbers has more than countably many members.⁶ So, since a language can include only countably many definitions, some of these members are undefinable. How might we, on Hansen’s behalf, respond to this argument?

One thing we might do is to appeal to Skolem’s Paradox, to point out that, by the Löwenheim–Skolem Theorem, the set-theoretic axioms that allow us to prove Cantor’s result have a countable model. So we might insist, following Skolem (1923) (see Bays (2014) for discussion), that our set-theoretic notions can only be pinned

⁵Compare Button and Walsh’s observation that ‘it is impossible to extract philosophical juice from a piece of pure mathematics, without invoking some philosophical thesis’ (2018: p. 156).

⁶Actually, Hansen tends to focus on the set of real numbers rather than on the power set of the set of natural numbers. But he accepts that, for the purpose of his discussion about set size, these two sets are interchangeable.

down by the axioms themselves, in which case we don't have any determinate notion of set or power set (or countability or uncountability) in the first place. But this response isn't available to Hansen. Elsewhere in the chapter, he asserts, for example, that he 'understand[s] the structure of an omega sequence' (p. 14) and that he 'think[s] the *idea* of the set of all possible combinatorial subsets of a given set x is so clear that that there is no wiggle room' for there to be multiple realizations of that idea (p. 27). So there's simply no way for him to insist that we haven't pinned down determinate notions here.

What he *does* do, though, is something fairly similar to this. He points out that Cantor's proof can be replicated in an alternative interpretation (in particular, the interpretation he goes on to develop later in the book) on which all sets and functions are describable: the assumption that there's a bijection between the set of natural numbers $\mathbb{N}$ and its power set $\mathcal{P}(\mathbb{N})$ implies, on this interpretation, that there's a description of that bijection, from which it will be possible to construct, by diagonalization, a description of some element of $\mathcal{P}(\mathbb{N})$ that isn't in the range of the bijection. But the language can only contain countably many descriptions in total, which means $\mathcal{P}(\mathbb{N})$ is a subset of a countable set.⁷ So it can't be that the nonexistence of the relevant bijection shows that $\mathcal{P}(\mathbb{N})$ has more members than $\mathbb{N}$ does. Hansen infers from the availability of this interpretation that Cantor's Theorem doesn't genuinely establish the existence of sets with more than countably many members.

But this is just to change the subject, by Hansen's own lights. For the fact remains that, on the *intended* interpretation—which, again, Hansen takes himself to have a grip on—Cantor's Theorem *does* tell us that $\mathcal{P}(\mathbb{N})$ has more members than $\mathbb{N}$ does. On that interpretation, a function from one set to another may be *any* mapping, describable or not, from members of the first set to members of the second. So what Cantor's proof shows is that the assumption that some such mapping is a bijection between $\mathbb{N}$ and $\mathcal{P}(\mathbb{N})$ leads to *contradiction*. This is a much stronger result

⁷This doesn't really make sense as it stands: if each member of $\mathcal{P}(\mathbb{N})$ corresponds to a description in a given language, it should be possible to generate a describable bijection between those expressions and the natural numbers simply by ordering the descriptions alphabetically. Hansen tries to respond to this sort of worry in Chapter 10 by appeal to the thought that no single fixed language can contain every possible member of $\mathcal{P}(\mathbb{N})$ —hence the need for an open-ended language. (Here he continues to insist that $\mathcal{P}(\mathbb{N})$ has no more members than $\mathbb{N}$ does, suggesting that a conclusion to the contrary depends on 'a quintessentially platonic belief that whenever two classes are of the same size (in an intuitive sense of "same size"), there must be a bijection between them, consisting—if everything else fails—of arbitrary ordered pairs of one element from each of the two classes,' and that, from a nominalist perspective, 'there is no reason to think that the absence of the latter, in the case of two infinite classes, tells us anything that can reasonably be described as a "difference in size"' (p. 228). I confess that I have no idea what 'intuitive' notion of sameness of size is supposed to be available by appeal to which we can avoid concluding here that $\mathcal{P}(\mathbb{N})$ is larger than $\mathbb{N}$.)

than Hansen's reinterpreted one: it tells us, not just that there's some metaphysical constraint ruling out the existence of a function *showing* that there's a unique member of $\mathbb{N}$ for each member of $\mathcal{P}(\mathbb{N})$, but that it's *logically impossible* for there to be a unique member of $\mathbb{N}$ for each member of $\mathcal{P}(\mathbb{N})$.⁸ It's not clear what more could be needed to establish that $\mathcal{P}(\mathbb{N})$ has more than countably many members.

Hansen, then, can't avoid Cantor's result simply by insisting on a reinterpretation of it. He must say instead that the problem is that, insofar as the existence of $\mathcal{P}(\mathbb{N})$ would require the existence of undefinable sets, this is reason to doubt that $\mathcal{P}(\mathbb{N})$ can exist at all, in which case the only lesson we should take from Cantor's Theorem is that ' $\mathcal{P}(\mathbb{N})$ ' fails to refer, not because we don't have any determinate notion corresponding to that term but because it turns out that no possible object answers to the notion we have. If we take for granted that to believe there are undefinable sets is to accept metaphysical platonism, this is reasonable enough. But now we're back to the uninteresting claim that mathematical results *within* set theory can't give us reason to accept metaphysical platonism *about* set theory, and it's unclear why there was any need to wade into the details of the mathematical results, interesting though they are.

What if we *don't* take this sort of platonistic implication for granted? Then we might ask what Cantor's Theorem tells us given (for instance) a broadly Carnapian deflationary approach to set-theoretic ontology. But such questions go unexplored in Hansen's discussion, since he has defined away the possibility of any combinatorialist attitude toward set theory that isn't paired with a commitment to metaphysical platonism about abstract objects.

We might worry that this last complaint is uncharitable. In Chapter 1 Hansen says that what he rejects on the basis of his skepticism about abstract objects is 'classical mathematics (understood non-formalistically)' (p. 4), and in Chapter 6 he suggests that a formalist is someone who 'takes stipulated axioms and inference rules to be basic' (p. 98). We might suspect, then, that on Hansen's way of taxonomizing the available views, those who accept a Carnapian sort of deflationism about set-theoretic ontology are to be characterized not as combinatorialists after all but instead as formalists, in which case Hansen's arguments against this sort of deflationism are to be found in his critique of formalism rather than in his critique of plenitudinous combinatorialism.

But this can't be right. The sort of formalist view Hansen discusses is one on which 'axioms are sentences that are stipulated to be true outright, and the stipulated truth condition for all other sentences is that a syntactic proof satisfying

⁸This is to say that Hansen's doubts about the possibility of undecidable functions don't give him a way to avoid this result, since the force of the result doesn't require the assumed bijection to be possible. It requires only that the assumption that there's an undecidable bijection doesn't on its own lead to contradiction.

certain conditions could exist' (p. 98). And this particular understanding of what formalism amounts to plays a central role in Hansen's discussion—his criticisms (e.g., that it's impossible, for Gödelian reasons, to reconstruct all of classical arithmetic formalistically) depend entirely on the idea that, on a formalist approach, taking on board a system of axioms and rules amounts to stipulating that a mathematical sentence is true just in case it's provable using those axioms and rules. This sort of stipulation, though, has no place in a Carnapian deflationary picture. On such a picture, taking on board a system of axioms and rules simply amounts to choosing to speak in a particular way, with truth conditions for the relevant sentences just given by some broadly Tarskian rules governing the truth predicate.⁹ So Hansen's critique of formalism doesn't tell us what's wrong with this sort of deflationism, any more than his critique of plenitudinous combinatorialism does.

This isn't to claim that there's nothing to be said against deflationary approaches to mathematical ontology, of course. The point is that Hansen himself doesn't discuss the tenability of such approaches and, more generally, doesn't engage with metaontological questions at all. Indeed, he doesn't appear even to acknowledge the existence of such questions, despite the fact that there has been significant debate about them in the recent philosophical literature. This is a serious omission. Hansen is recommending revising mathematical practice on philosophical grounds, and so he must show that this revision is genuinely necessary (as he recognizes, which is why he takes the critique in Chapter 2 to be needed to motivate his project). But showing this involves explaining why attempts to respect his metaphysical and epistemological scruples *without* revising mathematical practice—including attempts that depend on a deflationary metaontological picture—are bound to fail. And this he doesn't even begin to do.

The view Warren lays out in *Shadows of Syntax* can be understood as a particularly fully developed example of the sort of broadly Carnapian picture Hansen ignores. Warren starts from a radically permissive sort of inferentialist metasemantics, a view according to which we're disposed as a matter of convention to treat as binding certain syntactically specified rules governing the use of logical and mathematical expressions in inference—i.e., to enforce those rules, to comply with them in certain circumstances, and so on—and those rules, *whatever they are*, are meaning-constituting for the expressions in question and so are 'automatically valid' (p. 58). The automatic validity of those rules is, Warren says, the source of all logical and mathematical truth. The resulting picture is one on which logical and

⁹For Carnap himself, semantic predicates are just additional linguistic expressions, governed by their own rules—this is clear enough from his discussion, in Section 4 of 'Empiricism, Semantics, and Ontology' (1950), of statements like "Five" designates a number'. And if that's right, taking on board axioms and rules can't amount to stipulating truth conditions for sentences, on pain of circularity.

mathematical truths (along with numbers, sets, and other abstract mathematical objects) are simply ‘reflections of the syntactic rules of our language’ that ‘answer to nothing external save for our varying needs’, with ‘no deeper nature’ to be uncovered (p. 317).

Warren’s book has been much discussed in the few years since its publication, and deservedly so: it’s an ambitious book, full of (often convincing) arguments for theses entirely at odds with current philosophical orthodoxy. Indeed, every challenge to conventionalism or inferentialism that has ever received any attention in the literature is addressed somewhere in the book, directly and without any attempt at burden-shifting—Warren simply identifies the problems that need to be solved and then goes about trying to solve them.

The book consists of four parts of a few chapters each; I’ll give a (woefully incomplete) summary here. Part I contains important background material. In it, Warren clarifies that the conventionalist thesis he’ll be defending is an *explanatory* thesis to the effect that facts about logical and mathematical truth and necessity in a language are *fully explained* by the conventions of that language (more on this below); he lays out naturalistic constraints that disallow both ‘appeal[ing] to an independent realm of non-causal facts’ and ‘attribut[ing] to humans non-causal cognitive powers’ (p. 18); and, after explaining that the conventions that will be relevant for him are (for various reasons, some having to do with Quine’s regress argument) *not* explicit stipulations, he makes a case for *implicit* conventions by providing plausible functionalist–dispositionalist accounts of sentence acceptance, sentence rejection, inference, and inference rule-following. In performing this latter set of tasks, he points out that Quine-style skepticism about implicit rule-following is incompatible with the rule-based explanations of morphological and grammatical competence that are standard in linguistics and cognitive science, and he also provides a simple and extremely compelling defense of dispositionalism about rule-following against Kripkensteinian worries by appeal to what he calls ‘composite dispositions’ (p. 41)—i.e., dispositions, when in a situation in which one has completed a particular step of the process of (say) adding two very large numbers, to undertake the next step, so that one may have a composite disposition to complete *each* step of the process even if the sequence of steps is so long that one will never be able to complete *all* of them.

Part II is in my opinion the heart of the book. It’s here that Warren develops, for logical expressions in particular, the radically permissive—i.e., *unrestricted*—inferentialist view described above. On this view, certain rules of inference governing the use of a given expression are ‘meaning-determining’; in particular, they ‘determine the expression’s meaning by being automatically valid’, where validity amounts to ‘necessary truth preservation in the strongest possible sense’ and where for the rules’ validity to be automatic is for it to be the case that ‘given that they are meaning determining, nothing more is required of them for validity’ (pp. 58–9).

Furthermore—where ‘basic rules’ are just those that are ‘followed directly’, as ‘the foundational moves in our language game’, rather than those that are followed only ‘as a byproduct of directly following other rules’ (p. 66)—the rules that are meaning-determining and so automatically valid are just ‘the basic rules... whatever they are’ (p. 72). This is what makes Warren’s inferentialism *unrestricted* and what makes his view ‘a version of... conventionalism’ (p. 95). After clarifying all of this, Warren responds to various worries about unrestricted inferentialism and conventionalism, including but not limited to the classic ‘tonk’ problem, the worry that our inference rules don’t suffice to give determinate truth-conditional meanings to our logical expressions, and what he calls ‘the master argument’, according to which, since conventions can’t make *propositions* true and sentences are true just in case the propositions they express are true, conventions can’t make sentences true either. (I discuss his responses to the first two worries below; his response to the master argument is in my opinion not entirely satisfactory, for reasons I discuss in my ‘Linguistic Convention and Worldly Fact’ (2019) in reply to an earlier paper of Warren’s (his (2015)) in which he gives a version of that response.) He also gives straightforward accounts of what conventionalists should say about both logical disagreement and the epistemology of logic.

Part III extends to mathematical language the approach to logical language developed in Part II. Here Warren points out that we can conceive of the axioms of a given mathematical theory as meaning-determining zero-premise rules. He then suggests that the two main obstacles to a conventionalist approach to mathematics are, first, the worry that since many mathematical truths have existential import, the very idea that they can be true by convention is absurd, and second, the worry that, while it seems to be the case that ‘there is a *fact of the matter* even about unproved mathematical claims like Goldbach’s conjecture’ (p. 204), it’s not clear, given Gödel’s incompleteness theorems, how conventionalists can account for this sort of truth-theoretic determinacy.¹⁰ So he tries to fully answer both worries. I discuss his response to the second worry below; his response to the first worry just goes via a version of the broadly Carnapian sort of deflationary metaontology I discussed above—a metaontology on which, e.g., ‘the existence of numbers is not a substantive and controversial fact about our world, but is instead a *trivial byproduct* of our using our arithmetical language in the way that we do’ (p. 213). After giving these responses and clarifying some implications of his conventionalist view, Warren responds also to various technical metamathematical arguments against conventionalism, giving (in my opinion) compelling answers to all of them. And finally, he closes in Part IV with a set of meta-level reflections in which he clar-

¹⁰Warren acknowledges that it’s not obvious that *all* mathematical claims have determinate truth values, but he suggests that it’s ‘intuitively plausible’ that ‘the claims of basic arithmetic’, at least, are indeed determinate (p. 204).

ifies the relationships between his theory and some earlier conventionalist views, argues that the correctness of his theory is not itself a matter of convention but is instead a matter of genuine (nonconventional) metasemantic fact, and considers what he should think about his theory in light of the fact that conventionalism was once a popular view but is now almost universally rejected.

I am—cards on the table—extremely sympathetic to Warren’s project. The picture he develops is, in my view, almost certainly correct in its overall thrust. But even those who aren’t at all sympathetic will find a great deal of material here that’s worth reflecting on. For any objection to conventionalism that an unsympathetic reader takes to be fatal, it’s almost certain, as I’ve suggested, that the book contains a substantive response to that objection—a response that’s at least challenging if not entirely decisive. Warren says at one point that he takes himself to have ‘demonstrated that the arguments against conventionalism are either problematic on their own terms or aimed at implausible straw men’ (p. 332), and my estimation is that if this is an exaggeration, it’s only a slight one. Anyone interested in the foundations of logic and mathematics will benefit from careful engagement with the material in this book.

For all that, though, I must admit that I find myself a bit puzzled about some of the features of the conventionalist view Warren presents. There are, I take it, crucial matters of detail with respect to which it’s not obvious just how the view is to be understood, either because important implications have not been explicitly acknowledged or because the view itself has been formulated in a way that leaves its commitments less than fully clear. And this, it seems to me, leads in at least some cases to Warren misapprehending the dialectical situation he’s in, dismissing certain questions that he in fact owes answers to and, on the other hand, working to solve some problems that turn out, when the implications of his view are made clear, not to arise at all. I’ll close here by discussing some interrelated matters about which I think more clarity would be helpful, offering clarification myself where I feel able to do so.

One simple example of the sort of inexplicitness I have in mind is that, while Warren bills his unrestricted inferentialist view as a view about logical and mathematical language in particular, it in fact, as far as I can tell, commits him to unrestricted *global* inferentialism—i.e., the view that *all* linguistic expressions are governed by automatically valid conventional rules of inference, where the notion of inference is here broadly construed so as to include, in addition to intralanguage transitions of the usual sort, language entry and departure transitions of the sort described by, e.g., Sellars (1954). To see why, note that the correct response to the ‘tonk’ problem, from Warren’s unrestricted logical inferentialist perspective, is to insist that ‘tonk’ is a meaningful connective after all and that if we were to add the ‘tonk’ rules to our language, those rules would in fact be valid despite the fact that they would allow us to prove every sentence of the language: adding those rules

to our language would ‘drastically change[] the meaning of every sentence of the language’ (p. 93), with the resulting language—*Tonklisch*—being ‘a trivial language in which every sentence is true’ (p. 94).¹¹ Since we can use the ‘tonk’ rules to establish (to use one of Warren’s own examples) ‘Saul Kripke was born before Plato’, that sentence, despite being false in English, is a trivial truth in Tonklisch. To suppose otherwise—to suppose that ‘Saul Kripke was born before Plato’ has the same meaning in Tonklisch that it does in English and so is false in Tonklisch—is, Warren suggests, to make what he calls *the translation mistake*: ‘the mistake of being misled by superficial features into thinking that a homophonic translation is appropriate when it is not’ (p. 126).

What all of this implies is that the ‘tonk’ rules have a role to play in determining the meaning not just of ‘tonk’ itself but of *every* expression of the language. If this weren’t so—if the metasemantics of the expressions in sentences like ‘Saul Kripke was born before Plato’ were in some way essentially separable from the metasemantics of logical expressions—then adding the ‘tonk’ rules to our language could have no effect on the meaning of sentences like ‘Saul Kripke was born before Plato’. So the ‘translation mistake’, as applied to such sentences, wouldn’t be a mistake after all, in which case Warren’s defense of the claim that the ‘tonk’ rules are valid, and hence of *unrestricted* inferentialism, would fail for obvious reasons. And this is to say that the view Warren develops is and must be a view on which the meanings of *all* expressions, not just the logical and mathematical ones, are inferentially determined in such a way that the rules we follow turn out to be valid no matter what those rules are.

This isn’t explicit in Warren’s text—though he mentions global inferentialism a few times, he at least sometimes (e.g., on p. 131) seems to suggest that endorsing it is optional. But it’s clear enough, I take it, that in reality, it is straightforwardly a commitment of his view.¹² This is not an objection. As I’ve suggested, I’m very

¹¹ Recall that ‘tonk’ (see Prior (1960)) is an invented connective governed by an introduction rule allowing the inference from A to ‘ A tonk B ’ and an elimination rule allowing the inference from ‘ A tonk B ’ to B , for arbitrary sentences A and B . An application of the introduction rule followed by an application of the elimination rule would allow us to derive any sentence from any other—this is why adding ‘tonk’ to our language would render every sentence of the language provable.

¹² Strictly speaking, there may be logical space here for a view on which certain expressions can have meanings that aren’t inferentially determined but on which the expressions in question are necessarily blocked from having such meanings in all cases in which their having those meanings would invalidate the inference rules we *do* follow. And whether such a view counts as a species of global inferentialism is perhaps a verbal question. But it’s not clear what could motivate a view with this structure: if there are meanings that aren’t inferentially determined, why should it be the case that, when those meanings come into conflict with the rules of inference we follow, the meanings that aren’t inferentially determined are *always*, without exception, overridden by the rules of inference? The larger point, in any event, is that Warren’s unrestricted logical inferentialism commits him, in a way he doesn’t explicitly acknowledge, to a fairly radical view about the metasemantics

sympathetic to the view in question, and indeed to its global implications; I *agree* that the ‘tonk’ rules are valid in languages that include them, that this isn’t a genuine problem, and that to think otherwise on either count is to fail to appreciate fully the inferentialist perspective—see my ‘Whence Admissibility Constraints?’ (2026). The point is that global inferentialism—a view significantly further from the philosophical mainstream than is inferentialism about logical and mathematical language—is, from what I can tell, central to Warren’s picture in a way that goes unacknowledged in the book.

A second way in which it’s less than obvious what Warren’s commitments are seems to me to be more difficult to resolve. As briefly mentioned above, Warren describes his conventionalist view as one on which ‘facts about logical truth, logical falsity, logical necessity, and logical validity’ in our language, along with ‘facts about mathematical truth and falsity’ in our language, are *fully explained* by our linguistic conventions (p. 10), in ‘the worldly sense’ of explanation and not just in ‘the communicative sense’ (p. 12)—where the explanations in question ‘aren’t *causal* in a standard sense’ (p. 13)—but he insists at the same time that the notion of explanation he has in mind is not ‘a *metaphysically* robust notion’ like grounding or metaphysical dependence, that he ‘find[s] these notions of heavyweight metaphysics puzzling’, and that what’s needed here is just ‘the perfectly reasonable everyday notion of explanation’, which ‘is not a heavyweight metaphysical notion of the kind envisioned in recent philosophy’ (pp. 15–6). It’s clear enough here that Warren is trying to distance his view from metaphysicians’ claims about (e.g.) grounding, claims that he takes to violate naturalistic strictures.¹³ But whether he has successfully drawn any genuine distinction between ‘heavyweight’ and ‘lightweight’ notions that allows him to avoid commitment to the sort of metaphysical theorizing he takes to be problematic is less clear; simply insisting that he’s not doing heavyweight metaphysics doesn’t make it so.¹⁴ The explanatory claims that constitute his

of *all* linguistic expressions. (To be clear: Warren *is* explicit about the fact that adding ‘tonk’ to our language would have an effect on the meaning of every expression of the language. But what this tells us more generally about the metasemantics of those expressions isn’t discussed.)

¹³ Similarly for various other points in the book where he distinguishes between ‘metaphysically heavyweight’ and ‘metaphysically lightweight’ readings of a claim and then insists that he’s not committed to the heavyweight version: in each case the goal is to distance his view from the sort of metaphysics he finds naturalistically unacceptable.

¹⁴ He suggests that grounding and related notions are notions of a ‘relation [that] is meant to be an objective, factual, hyperintensional relationship *that is somehow metaphysical in character*’ (p. 16, my emphasis), as though the ‘metaphysical character’ of the relation is supposed to be something more than its objectiveness or its factuality, in which case he can help himself to notion of a worldly, objective, factual explanatory relation that is nevertheless *not* metaphysical. But it’s not at all clear what more could be required for the relation in question to be ‘metaphysical’ than that it be worldly, objective, factual, and noncausal. And indeed, according to at least some grounding theorists—e.g., Audi (2012), who isn’t taking any sort of deflationary attitude toward the notion and who is surely

theory are, again, ‘worldly’ claims about how things really are—as he makes clear much later in the book, his view is that ‘metasemantics is a matter of fact’ (p. 337) and that his ‘conventionalist theory of logic and mathematics is *factually true* [and] *uniquely correct*’ (p. 341). It appears, then, that the explanatory claims in question are to be understood as substantive claims about what is objectively the case in reality. So in what sense are they merely ‘lightweight’? I’m not sure what to make of this.

We might suppose that we can clarify matters here by returning to the particulars of Warren’s unrestricted inferentialist theory, in order to determine what sorts of explanatory claims that theory entails. So: inferentialism, as Warren describes it, is a view on which the rules for an expression ‘explain what the expression means, directly, and without any explanatory contribution from other sources’ (p. 57), where, in the case of logical expressions, the meanings in question are *truth-conditional* meanings; this explanatory claim, he suggests, is just equivalent to the claim that the rules are ‘meaning-determining’ or ‘meaning-constituting’. And the picture here, recall, is one on which, *whatever rules we follow*, those rules are *guaranteed* to be valid, since their validity is automatic in that it requires nothing more than that we follow them. To say that those rules explain what some of the expressions of our language mean, then, is, in this context, to say that our following those rules suffices on its own for the expressions in question to have truth-conditional meanings on which the rules are necessarily truth-preserving. This, it seems to me, *confirms* that claims involving the explanation relation that’s in play here are substantive claims about what the world is like.¹⁵ If there’s a sense in which that relation is nevertheless metaphysically lightweight, I haven’t been able to determine what

doing heavyweight metaphysics if anyone is—grounding just *is* whatever determination relation underpins noncausal worldly explanations; this, not some further mysterious quality, is what its metaphysical character consists in.

¹⁵At one point Warren describes the ‘logical facts’ that are ‘the target of conventionalist explanations’ as ‘lightweight’ facts that, ‘even if “about” the world in a thin sense, are utterly empty’ (p. 112). If this could be sustained, it would perhaps clarify the situation: insofar as the facts to be explained don’t really tell us anything about what the world is like, explanations of those facts presumably aren’t going to tell us anything substantive about what the world is like either. But it can’t be sustained. The facts that our conventions are supposed to explain, as Warren makes clear, are metalinguistic facts such as the fact that “either it is raining or it is not raining” is true in our language’ (p. 10), and these are certainly not empty: they’re contingent facts about the language we speak. (Warren, just before describing the facts that our conventions are supposed to explain as empty, explains that on the ‘lightweight’ notion of a fact, ‘it is a fact that *p* just in case “*p*” is true’, in which case ‘it is a fact that either the Red Sox are the best team in baseball or the Red Sox are not the best team in baseball’ (p. 112). *This* fact may well be empty, but it is an object-level logical fact, not a metalinguistic fact about what truth value a sentence has in a particular natural language; object-level facts are not the primary target of his conventionalist explanations. I suspect that Warren doesn’t clearly distinguish between the two types of facts in this context for reasons having to do with his deflationism, about which more below.)

that sense is.

The claim that our rules are automatically valid, though, leads to a further question: if it's genuinely the case that any rules we follow are thereby guaranteed—with nothing further required—to be truth-preserving, what can the truth of the sentences that appear in inferences sanctioned by those rules amount to? That is, what is the correct approach to truth and related semantic notions, from a conventionalist perspective? Warren doesn't say all that much about this question, and what he does tell us is less than fully committal. He says in one footnote that conventionalism and deflationism 'neatly cohere' to such a degree that 'a combined view—inferentialist deflationism—is, I think, the best overall approach to matters of meaning' (p. 111fn), but he also says, in a later footnote, that 'there is a lot of wiggle room' with respect to 'the question of which theory of truth conventionalism is committed to' and that 'only explanatorily robust versions of the correspondence theory are ruled out' (p. 115fn). It seems to me, though, that endorsing some sort of deflationary theory is in fact the *only* option. To see why, note that the truth-conditional meanings that logical expressions have as a result of our following the rules we do certainly can't be worldly entities of any kind, if nothing more is to be required, beyond our following those rules, in order for the expressions to have those meanings: if they're worldly entities, they can be the meanings of our expressions only if the world in fact contains them, which is a requirement beyond our following the rules.

Another way of seeing the point is to note that, given global inferentialism, everything there is to know about semantic properties is given by the inference rules governing our semantic predicates, where those rules aren't going to resemble rules governing empirical predicates but instead are going to include rules that look more like the usual disquotational truth rules; Warren himself (p. 153) discusses such rules. This, as far as I can tell, just *entails* that the correct theory of these semantic notions is a deflationary theory of some kind.¹⁶

It's worth acknowledging that the issue here may be a merely verbal one: perhaps Warren is simply working with a narrower conception than I am of what it is for a theory of semantic notions to be deflationary, a conception on which a theory may be nondeflationary even if it's a theory on which semantic predicates are

¹⁶If the truth predicate is governed by purely disquotational rules, it may be that, insofar as the fact that either it is raining or it is not raining is empty, the fact that 'either it is raining or it is not raining' is true is empty as well. This, despite appearances, doesn't contradict my previous footnote. A purely disquotational truth predicate can be applied only *within* a language, to interpreted sentences of that very language. To claim that a sentence is *true in a particular language*, on the other hand, is to apply a truth predicate to a syntactically individuated sentence that may appear in more than one language, where for the sentence to fall under the truth predicate requires that it have a particular meaning in the language in question; this is a substantive requirement. (This latter truth predicate may still be disquotational in an extended sense—see Field (1994).)

simply more words of our language that are governed by their own rules and that aren't associated with worldly semantic entities. But in any event, the fact that semantic predicates, on Warren's view, do have these features—whether the theory of semantic notions to which that fact commits him counts as deflationary or not—seems to me to have metasemantic implications that he doesn't recognize. Warren introduces, as a genuine metasemantic worry about his inferentialism, Carnap's so-called *Categoricity Problem* (see his (1943)), which, in this context, amounts to the fact that the validity of the rules of standard natural deduction calculi is compatible with deviant interpretations of the propositional constants as well as of the first-order quantifiers, in which case the validity of those rules isn't sufficient to pin down the standard interpretation of those expressions. (One simple example of the problem: the validity of the rules is compatible with its being the case that every sentence of the language is a necessary truth, since in that case no argument sanctioned by the rules will ever have a false conclusion. But if every sentence of the language is a necessary truth, the negation sign certainly doesn't have its standard meaning, since for any sentence, both it and its negation are guaranteed to be true.) The fact that bilateral rules, which govern sentence rejection as well as sentence acceptance, are known to make available a solution to this problem in the propositional case appears to be one of Warren's reasons for insisting on bilateralism: he describes 'the ability to cleanly solve Carnap's problem' as 'perhaps [bilateralism's] most important advantage' (p. 81). And similarly, in discussing the first-order quantifiers he emphasizes that logical rules are to be 'understood in a fully open-ended fashion' (p. 85)—i.e., as valid not just in our own language but also in expansions of that language—since this suffices, under certain assumptions, to solve the problem in the first-order case as well. But working to solve Carnap's problem here is, from what I can tell, just a mistake: Warren's commitments with respect to semantic predicates turn out to entail that the problem can't arise in the first place. That is, there can be no genuine metasemantic worry of this kind, on Warren's inferentialist picture.

To see why not, note that if our rules include disquotational truth rules along with the standard propositional natural deduction rules, those rules are trivially going to allow us to prove, for any sentence of our language, the claim that the sentence's negation is true if and only if the sentence isn't true.¹⁷ Or, even more directly: if our rules include the clauses of a standard Tarskian truth-definition, they will just

¹⁷If we start by assuming both that '*S*' is true and that 'it's not the case that *S*' is true, we can derive a contradiction just by applying the truth predicate's elimination rule to both assumptions, which means we can conclude, by reductio and conditional proof, both that if '*S*' is true then it's not the case that 'It's not the case that *S*' is true and that if 'it's not the case that *S*' is true then it's not the case that '*S*' is true. Note that, though we can prove this for *any* sentence of the language, we can't prove the corresponding universal claim—i.e., that it holds for *all* sentences of the language—unless we have access to substitutional quantification or some related device.

include, as an axiom, the claim that any sentence's negation is true if and only if that sentence isn't true. (And analogously for the rest of the logical constants.) In both cases, the claims in question will be provable using the rules of our language and so will be guaranteed to be true, given Warren's inferentialism. So there's no genuine possibility of our logical expressions failing to have their standard meanings that needs to be ruled out by bilateralism, open-endedness, or anything else.¹⁸

It might be supposed that I'm simply confusing object language with metalanguage here. But this is not so. Warren's theory, as he makes clear repeatedly, is a theory of natural language, 'a philosophical theory about logic and mathematics *in the wild*' (p. 11). And in a context in which the language under discussion is our home language, there isn't a separate, antecedently available metalanguage for us to move to. So insofar as we're working in a metalanguage, it's just another portion of our natural language, governed by its own rules—including bridge principles connecting it, *in whatever way we as a linguistic community have landed upon*, to the portion of natural language we're theorizing about. Or, to put the point another way: there's no reason to take the natural deduction rules governing 'object-language' inferences to be the only meaning-constituting rules that are relevant in this context; rules governing inferences in the 'metalanguage' can surely play a role as well. Truth-conditional semantics, in this setting, is ultimately just more syntax. This is another way of seeing that there's no genuine sense in which the validity of the rules of our language might turn out to be compatible with our logical constants having been given the 'wrong' truth-conditional meanings. (Or at least, if this is a genuine possibility, I haven't been able to make sense of how.)

Something similar holds, I take it, for the broadly analogous determinacy worries Warren takes to arise in the mathematical case. As mentioned above, he takes 'accounting for the determinacy of basic mathematical truth'—in particular, the determinacy of *arithmetical* truth—in the face of Gödel's incompleteness theorems to be a 'major obstacle for mathematical conventionalism' (p. 239), and he also suggests that there are reasons for concern about whether our theory of arithmetic is *categorical*—i.e., whether we can rule out 'nonstandard interpretations of arithmetic' (p. 271). So he spends a chapter considering these problems, arguing in the end that another appeal to open-endedness allows us to rule out nonstandard models and that, in order to account for the determinacy of arithmetical truth, we can insist that, despite our finitude, we can and do follow the omega rule: closing even

¹⁸This is a version of Putnam's observation (in a discussion of names and predicates rather than of logical expressions) that from the perspective of "nonrealist" semantics—on which understanding doesn't consist in knowledge of truth conditions but on which it's nevertheless not necessary to 'fore swear forever the notion of a model', since it's straightforward to introduce model-theoretic talk into our language via the usual Tarskian definitions—the possibility of unintended interpretations simply doesn't arise, since 'the "gap" between words and world, between our *use* of the language and its "objects", never appears' (1980: pp. 479–80).

a weak arithmetical theory like Robinson arithmetic under the omega rule results in a complete theory. Warren's argument for the claim that we can follow infinitary rules like the omega rule is interesting and clever, though it seems clear to me that it ultimately fails, for reasons I discuss elsewhere—see my 'Can We Follow the Omega Rule?' ([forthcoming](#)). But here I just want to point out that, as far as I can tell, this case is like the case of the logical constants in that there isn't any genuine metasemantic problem to be solved.

The question of arithmetical categoricity doesn't arise for reasons directly analogous to the reasons Carnap's problem doesn't arise. We have a notion, which we can state metalinguistically, of what it is for a model of arithmetic to be standard. We're willing to say, for instance, that a model is standard just in case it has the structure of an omega sequence or, equivalently, just in case it's isomorphic to the smallest successor-closed sequence of von Neumann ordinals that begins with $\emptyset$. And when a model does *not* have this structure, we say that it's not the model we intend. (These claims may be basic, or they may be derived from other metalinguistic principles. It doesn't matter which.) So, since principles governing what we say in the 'metalanguage' can be meaning-constituting, there just is no chance of our arithmetical expressions failing to be correctly interpreted via the standard model.

As for the question of whether the *truth values* of arithmetical sentences are determinate: insofar as this sort of determinacy is itself understood model-theoretically, as sameness of truth value 'in each admissible interpretation' (p. 207), there's a standard proof that determinacy follows from categoricity, as Warren himself notes (p. 261). But he expresses 'a similar worry at the meta-level' (p. 262), to the effect that models of arithmetic may *themselves* be indeterminate, in which case proving that a given sentence has the same truth value in all admissible interpretations doesn't suffice to show that it has a determinate truth value after all. If determinacy is understood model-theoretically, though, there's no room for such a worry: the proof from categoricity to determinacy shows conclusively that the sentences of arithmetic are determinate, by the definition of determinacy.¹⁹

This suggests that, for Warren, determinacy is *not* to be understood model-theoretically. But he isn't explicit about this, nor does he make clear how determinacy is to be understood instead. He may well have in mind some straightforward conception of determinacy on which there's a substantive problem to be solved, but

¹⁹Note also that determinacy, as applied to *models*, is not the same property as arithmetical determinacy: it's not a metalinguistically specified property that arithmetical sentences might have or lack but a metametalinguistically specified property that model-theoretic sentences might have or lack. In order for the question of the determinacy of models even to arise, then, we must be able to interpret model-theoretic talk in a metametalanguage. And I suspect that if we can do this, then we can, by reasoning analogous to that outlined above, rule out nonstandard interpretations of our model theory and then use this categoricity result to demonstrate determinacy.

if so, he never articulates that conception.²⁰ The only actual definition of determinacy he ever offers is the model-theoretic one. And on that definition, there just can be no genuine problem here.

Funding

This work was funded by the Austrian Science Fund (FWF) [10.55776/P33708]; and by the Italian Ministry of University and Research (MUR) under Ministerial Decree (D.D.) No. 1236 of 1 August 2023 through the FIS 2 Call, under the project Humanizing Mathematical Knowledge: Fallibility, Technology, Know-How (Project No. FIS-04053; PI: De Toffoli; CUP I53C24003170001).

Acknowledgments

Thanks to Casper Storm Hansen and Jared Warren for their comments on an earlier draft of this review. Thanks also to Julien Murzi along with the participants in reading groups at Brown University, the University of Salzburg, and IUSS Pavia.

References

- Audi, Paul [2012]: ‘Grounding: Toward a theory of the *in-virtue-of* relation’, *Journal of Philosophy* 109, 685–711.
- Bays, Timothy [2014]: ‘Skolem’s Paradox’, in E. N. Zalta & U. Nodelman, eds., *The Stanford Encyclopedia of Philosophy*, Winter 2014 ed. Metaphysics Research Lab, Stanford University.
- Button, Tim and Walsh, Sean [2018]: *Philosophy and Model Theory*. Oxford University Press.
- Carnap, Rudolf [1943]: *Formalization of Logic*. Harvard University Press.
- Carnap, Rudolf [1950]: ‘Empiricism, semantics, and ontology’, in *Meaning and Necessity: A Study in Semantics and Modal Logic*, 2nd ed., pp. 205–21. University of Chicago Press, 1956.
- Field, Hartry [1994]: ‘Deflationist views of meaning and content’, *Mind* 103, 249–85.

²⁰It may be—though Warren never explicitly says anything like this—that the conception of determinacy he’s working with is one on which for arithmetic to be determinate is for every arithmetical sentence to be such that the rules of our language either *determine that it’s true* or *determine that it’s false*. (Note that on this conception, it’s not enough for the rules of the language only to *determine that the sentence is either true or false*. This makes available an explanation of why bivalence doesn’t suffice to secure determinacy that, unlike Warren’s own explanation (see p. 207), *doesn’t* go via the model-theoretic definition of determinacy.) But on this conception, there doesn’t seem to be any real gap between incompleteness and indeterminacy. So it’s not clear to me why indeterminacy, on this conception, is supposed to be particularly surprising or unsettling.

- Kripke, Saul A. [2024]: 'The question of logic', *Mind* 133, 1–36.
- Padro (now Birman), Romina [2015]: *What the Tortoise Said to Kripke: The Adoption Problem and the Epistemology of Logic*. Ph.D. thesis, CUNY Graduate Center.
- Prior, A. N. [1960]: 'The runabout inference-ticket', *Analysis* 21, 38–9.
- Purnam, Hilary [1980]: 'Models and reality', *Journal of Symbolic Logic* 45, 464–82.
- Quine, Willard V. [1936]: 'Truth by convention', In O. H. Lee, ed., *Philosophical Essays for Alfred North Whitehead*, pp. 90–124. Longmans, Green and Co.
- Sellars, Wilfrid [1954]: 'Some reflections on language games', *Philosophy of Science* 21, 204–28.
- Skolem, Thoralf [1923]: 'Some remarks on axiomatized set theory', in J. van Heijenoort, ed., *From Frege to Gödel: A Source Book in Mathematical Logic, 1897–1931*, pp. 290–301. Harvard University Press, 1964.
- Topey, Brett [2019]: 'Linguistic convention and worldly fact: Prospects for a naturalist theory of the a priori', *Philosophical Studies* 176, 1725–52.
- Topey, Brett [2026]: 'Whence admissibility constraints? From inferentialism to tolerance', *Journal of Philosophical Logic* 55, 411–35.
- Topey, Brett [forthcoming]: 'Can we follow the omega rule?', *Philosophical Quarterly*. DOI: <https://doi.org/10.1093/pq/pqafo59>.
- Warren, Jared [2015]: 'The possibility of truth by convention', *Philosophical Quarterly* 65, 84–93.