Internalism and the Determinacy of Mathematics

Lavinia Picollo
National University of Singapore, Singapore.
l.picollo@nus.edu.sg

Daniel Waxman
National University of Singapore, Singapore.
danielwaxman@nus.edu.sg

A major challenge in the philosophy of mathematics is to explain how mathematical language can pick out unique structures and acquire determinate content. In recent work, Button and Walsh have introduced a view they call 'internalism', according to which mathematical content is explained by internal categoricity results formulated and proven in second-order logic. In this paper, we critically examine the internalist response to the challenge and discuss the philosophical significance of internal categoricity results. Surprisingly, as we argue, while internalism arguably explains how we pick out unique mathematical structures, this does not suffice to account for the determinacy of mathematical discourse.

1. Introduction

Many philosophers have been concerned with the challenge, made famous by Benacerraf (1973), of providing a naturalistic account of the epistemology of mathematics. But there is a perhaps more fundamental challenge, raised originally by Putnam (1980), of providing a naturalistic account of the *content* of mathematical language.[1]

According to a pre-theoretically appealing picture of mathematical language and its content, large and significant areas of mathematics have the following features:[2]

[1] All of the issues we will discuss apply equally well to mental as to linguistic content. However, for brevity, we will speak as if the problem is fundamentally a linguistic one.

[2] This picture is not obviously committed to platonism or ontological realism about mathematics. Its appeal to a notion of *structure* can be understood for now loosely, used in roughly the informal mathematical sense. Although some theoretical philosophical developments involve reifying structures as self-subsistent entities, nothing in the picture itself requires this. Indeed, the picture can be endorsed even by fictionalists, provided they regard the fiction as rich enough to settle every arithmetical question. (On fictionalism, see Field 1980 and the essays in Field 1989. Field's own views on determinacy are complex; see Field 1998 for an argument that arithmetical vocabulary is determinate in virtue of its inferential links to physical vocabulary and certain cosmological hypotheses.)

Mind, Vol. 132 . 528 . October 2023
https://doi.org/10.1093/mind/fzac073

> **Uniqueness.** The content of the relevant vocabulary is sufficiently rich to 'pick out' a *unique* mathematical structure.

> **Determinacy.** The content of the relevant vocabulary is sufficiently precise to ensure that no questions are left open, that is, each claim is either determinately true or determinately false.

Consider arithmetic. When we use the language of arithmetic, it is natural to suppose that we have a unique structure in mind: *the natural number structure*. As a consequence, it is natural to regard any arithmetical question as yielding a definite answer. Are there infinitely many prime numbers that differ by 2? The twin prime conjecture famously answers 'yes'. Although we have not currently proved or refuted the twin prime conjecture, it is intuitively hard to see how it could lack a determinate truth-value. After all, if we could run through the natural numbers, we would ultimately find out the answer. Of course, there are infinitely many numbers, so we cannot actually run through them; but this looks like a medical impossibility, not a mathematical one (cf. Russell 1935, p. 143).

So we face the following challenge:

> **Putnam's Challenge.** What, if anything, explains uniqueness and determinacy to the extent they arise in mathematics?[3]

One possible view is that we possess a special faculty—for example, 'mathematical intuition'—which somehow connects us with the mathematical realm and which helps to fix mathematical content, perhaps as perceptual or causal acquaintance helps to fix certain kinds of empirical content. However, it is difficult to reconcile this idea with naturalistic views of the human mind, which neither postulate the existence of such faculties nor hold out any real hope of vindicating them. Following Putnam, say that a view is *moderate* if it adheres to this kind of naturalism.

Moderation severely limits the resources available to explain mathematical content. If uniqueness and determinacy hold, then these metasemantic facts must presumably be explained by mathematical practice, that is, the way in which the relevant vocabulary is *used*.[4] More

[3] The *locus classicus* is Putnam (1980), although he does not put it in exactly these terms. A closely related ('metasemantic') explanatory challenge, focusing on determinacy, is developed by Warren and Waxman (2020). Button and Walsh (2018) present what they call a 'doxological' challenge focusing on uniqueness; their challenge is presented primarily as a problem for the view they call 'modelism', for more on which see §2.

[4] This claim might be denied by those who endorse a Lewis-inspired account of metasemantics assigning a central role to *eligibility*. Proponents of this view might hold that facts about the objective eligibility of certain mathematical structures partially explain the content of our thought and talk.

specifically, our mathematical practice can be characterized in terms of (a) the mathematical theories we accept, and (b) the logical resources used in drawing out the implications of our theories via proof. For now we don't assume any particular formalization of mathematical theories or logic — something much discussed in what follows. With that said, moderation *does* impose constraints on how exactly these notions are fleshed out. On naturalistic assumptions about creatures like us, any theory we are in a position to accept is recursively axiomatized, and every proof system we are in a position to use is one in which proofs are finite and the property of being a proof is algorithmically checkable.[5]

Putnam's Challenge is especially pressing for moderates in light of landmark results in the foundations of mathematics in the early twentieth century — in particular, the Löwenheim-Skolem theorems and Gödelian incompleteness — which provide strong prima facie obstacles for any naturalist account. In their recent book *Philosophy and Model Theory* and in stand-alone papers, Tim Button and Sean Walsh have developed a view which they call 'internalism' or 'internal realism' (reflecting its inspiration by the work of Putnam).[6] Button and Walsh motivate internalism primarily as a *moderate* account of mathematical content capable of answering Putnam's Challenge and overcoming the obstacles raised by the limitative metamathematical results above.

In what follows, we will critically examine internalism and the internalist response to Putnam's Challenge. We will argue that although internalism sheds considerable light on metasemantic questions surrounding mathematics, it is nevertheless unable to give a satisfying account of the determinacy of mathematical content. Ultimately, we will argue, the philosophical constraints imposed by the Löwenheim-Skolem theorems and Gödelian incompleteness — precisely the results

We put such views aside for two reasons. First, we have our doubts that they are really acceptable on naturalistic grounds. Second, the appeal to eligibility arguably requires ontological commitments to reified structures in order to get off the ground; consequently, it is worth considering how less committal views fare.

[5] The main source of support for the recursive axiomatizability of our theories comes from the computational theory of mind: in brief, if the human mind does not have computational powers that exceed those of any Turing machine, then the recursive axiomatizability of our theories follows. Of course, this theory of mind has been denied; see for instance Lucas (1961), Penrose (1989), and those following them. But our point here is simply that the computational view is plausibly a non-negotiable commitment of naturalism of the sort that moderates endorse.

[6] See Button and Walsh (2018) for the fullest expression of the view, and Button and Walsh (2016) and Button (2022). Button and Walsh (2018) are explicit that they should not be read as *advocating* internalism; their stance towards it is perhaps better described as one of sympathetic exploration (though see Button 2022 for a fuller defence). In order to avoid constant hedging of this kind, however, we will speak as if Button and Walsh straightforwardly defend the view.

responsible for generating the challenge in the first place — cannot be so easily overcome. While internalism is the main target of the paper, we discuss a number of important topics along the way, including the philosophical significance of (both internal and model-theoretic) categoricity arguments, the standing of second-order logic, the metasemantics of determinacy, and the relationship between uniqueness of structure and determinacy of content. Surprisingly, as we will argue, even if internalism successfully explains how we pick out unique mathematical structures, this does not suffice to account for the determinacy of the relevant mathematical discourse.

Before introducing internalism (§3) and our objections to it (§§4–5), we first say more about Putnam's Challenge and the failure of some natural responses to it.

2. Moderate modelism

Putnam's Challenge arises, fundamentally, from philosophical implications of the Löwenheim-Skolem theorems and Gödelian incompleteness. In introducing the difficulties these results pose for moderate views, we will first show how they arise given two highly substantive assumptions, both of which will later be relaxed. Our discussion focuses (for the most part) on arithmetic, but it should be clear how it generalizes.

The first assumption, widespread in the philosophy of mathematics, is that mathematical structures are properly understood in terms of isomorphism types, that is, classes of isomorphic *models*. Button and Walsh refer to this view as 'modelism'. Thus, in order to explain uniqueness and determinacy for arithmetic, moderate modelists must identify an arithmetical theory all of whose models are isomorphic to one another. The second assumption is that our theories can be properly formalized in *first-order* logic.

Given these two assumptions, trouble arises. Take uniqueness first. By the Löwenheim-Skolem theorems, any consistent first-order theory of the natural numbers (indeed, any theory with infinite models) has non-isomorphic models. Given modelism, this implies that no first-order theory of arithmetic is capable of picking out a unique structure. Turn next to determinacy. As Gödel showed, any consistent recursively axiomatized first-order theory enabling minimal arithmetical reasoning is bound to leave many statements undecided. So, given moderation, and if sentences are determinate to the extent that they are decided by the theories that capture our mathematical practice, this implies that no such theory can underwrite the determinacy of arithmetic.

The natural response is for moderate modelists to question the assumption that our theories are adequately formalized in first-order logic. A seemingly more promising strategy involves considering *second-order* formalizations instead, allowing quantification into predicate and function position.[7]

Second-order arithmetic, PA_2, is formulated in a second-order language, $\mathcal{L}_{\mathrm{PA}_2}$. All first-order axioms are retained except for the induction schema, which is replaced by a single second-order universally quantified axiom. Thus PA_2 can be identified with the following conjunction:[8]

$$(\mathrm{PA}_2) \quad \begin{aligned} &\mathrm{N}0 \wedge \forall x\,(\mathrm{N}x \to \mathrm{NS}x) \wedge \forall x\,(\mathrm{N}x \to \mathrm{S}\,x \neq 0) \wedge \\ &\forall x \forall y\,(\mathrm{N}x \wedge \mathrm{N}y \to (\mathrm{S}x = \mathrm{S}y \to x = y)) \quad \wedge \\ &\forall X\,(X0 \wedge \forall x\,(\mathrm{N}x \to (Xx \to X\mathrm{S}x)) \to \forall x\,(\mathrm{N}x \to Xx)) \end{aligned}$$

The interesting feature of second-order theories is that they appear to offer a way around the limitative results. In particular, the Löwenheim-Skolem theorems do not extend to second-order logic, provided that second-order consequence is defined in terms of *full* models. These are models in which the second-order quantifiers range over the full power set of the first-order domain. Indeed, relative to this class of models, second-order arithmetic is categorical (see Dedekind [1888] 1996):

Theorem 1 (Dedekind Categoricity). All full models of PA_2 are isomorphic.

Categoricity is a highly appealing feature. On the modelist view, categorical theories pin down a unique structure, since they have only models of a single isomorphism type. Categoricity also promises determinacy: if a theory is categorical, then for every sentence in the language of the theory, either it is true in every model of the axioms or its negation is:

Corollary 2 (Dedekind Intolerance). Every sentence of $\mathcal{L}_{\mathrm{PA}_2}$ is either true in all full models of PA_2 or false in all of them.

So, assuming the moderate modelist can appeal to full second-order logical consequence, a satisfying answer to Putnam's Challenge seems available. However, this appeal is usually regarded as unsuccessful, for reasons we broadly endorse.[9]

[7] For arguments that second-order languages better capture various aspects of informal mathematics, see Kreisel (1967) and Shapiro (1991).

[8] Button and Walsh refer to this theory as $\mathrm{PA}_{\mathrm{int}}$.

[9] For the *locus classicus*, see Weston (1976); for discussion, see also Field (2001, postscript to ch. 12), Parsons (2008, §48), and Warren (2020, p. 244).

Full models are a subclass of the broader class of (faithful) *Henkin* models, in which the second-order quantifiers may range over *any given subset* of the power set of the first-order domain that validates the Comprehension Schema.[10] Full models and Henkin models give rise to different notions of semantic consequence, corresponding to truth preservation in the relevant class of models. The differences are significant. There are recursive proof systems — natural generalizations of the familiar calculi for first-order logic — which are sound *and complete* with respect to Henkin consequence. On the other hand, it follows from Gödel's theorems that no recursive calculus can be sound and complete with respect to full consequence. Henkin consequence behaves much more like first-order consequence than full consequence. In particular, the Löwenheim-Skolem theorems apply, meaning that any theory with infinite models has non-isomorphic Henkin models; and, by completeness, Gödel's theorems entail that any theory containing enough arithmetic will give rise to sentences such that neither they nor their negations are true in all Henkin models of the axioms.

These technical details give rise to a powerful philosophical objection to the moderate modelist who wishes to appeal to Dedekind Categoricity and Intolerance. Recall that the 'moderate' part of moderate modelism involves a commitment to a naturalistic metasemantics; as we argued above, this involves explaining mathematical content in terms of (a) recursively axiomatized mathematical theories, and (b) a recursive notion of proof. But these resources simply cannot do what the moderate modelist needs: to exclude Henkin models as somehow unfaithful to our practice. For (if consistent) PA_2 has non-isomorphic Henkin models which make different sentences true. But no (recursive) calculus that is sound with respect to full consequence will be able to rule such models out. So moderate modelists cannot appeal to the notion of full second-order consequence to extract philosophical conclusions from Dedekind Categoricity and Intolerance.

The moderate modelist might argue that non-full models are excluded, not by PA_2 or its proof-theoretic consequences, but rather by ascending to a set-theoretic metatheory in which the distinction between full and Henkin models can be expressed. That is, the moderate modelist might hope to be able to define the notion of a full model of $\mathcal{L}_{PA_2}$ using set-theoretic vocabulary in such a way that, in every model of set theory, the objects satisfying the definition are indeed (all) full

[10] That is, $\exists X^n \forall x_1 \ldots \forall x_n \, (X^n x_1 \ldots x_n \leftrightarrow \phi(x_1, \ldots, x_n))$, where ϕ is any formula in which X^n does not occur free. For details see Shapiro (1991, §3.2).

models of $\mathcal{L}_{PA_2}$. The problem is simply that a directly analogous issue recurs at the level of the metatheory (see Parsons 2008, p. 274). (One might say, following Putnam 1980, that the metatheoretic distinction between full and Henkin models is 'just more theory'.) Suppose (for a moment) that the set-theoretic metatheory is a first-order theory. Then that theory will itself have non-isomorphic interpretations — $\mathcal{M}_1$ and $\mathcal{M}_2$ — with non-isomorphic internal models of PA_2 — ω_1 and ω_2: that is, ω_1 satisfies the definition of a full model of PA_2 in $\mathcal{M}_1$, and the same for ω_2 in $\mathcal{M}_2$. As long as 'non-standard' models cannot be ruled out for the metatheory, it cannot rule them out for the object theory either.

The moderate modelist will likely reply that these results are fatal *if the set-theoretic metatheory is a first-order theory*; but note that results analogous to Dedekind Categoricity and Intolerance can be proven for *second-order set theory*, ZFC_2.[11] However, here again the dialectic recurs. To argue that these results have philosophical significance, the moderate modelist must insist that only full models are admitted. For if Henkin models are allowed, then a version of Parsons's point holds once again: there are non-isomorphic Henkin models of ZFC_2 and, what is more, models with non-isomorphic *internally full* models of PA_2.[12] But this defeats the whole purpose of the move to a set-theoretic metatheory. The point of that move was to legitimize the restriction to full models; but the results above show that such a restriction must instead be *presupposed*.

In short: the moderate modelist simply lacks the metasemantic resources to exclude Henkin models, and thus cannot appeal to Dedekind Categoricity and Intolerance to explain uniqueness and determinacy.

3. Internalism

Button and Walsh's internalist alternative to modelism involves two central ideas. The first is a rejection of the thought that the apparatus of second-order logic requires any further *interpretation*, via set

[11] More precisely, ZFC_2 is *quasi*-categorical (as shown by Zermelo 1930): any two full models of ZFC_2 are either isomorphic to one another or are such that one is isomorphic to a proper initial segment of the other. There is also an analogue of intolerance: while quasi-categoricity does not enable full intolerance, it nevertheless holds for sentences which are 'decided' at low levels of the iterative hierarchy, such as the continuum hypothesis (which concerns the size of the continuum and is thus 'settled' in $V_{\omega+2}$); Kreisel (1967) famously argues from this to the determinacy of the continuum hypothesis. Of course, ZFC_2 is not the only second-order set theory in town: Button and Walsh (2018) propose a variant of Scott-Potter set theory for which they prove full, not merely quasi, categoricity and intolerance.

[12] 'Internally full' in the sense that they satisfy the definition of fullness in the ZFC_2 models at issue.

theory or otherwise. Rather, the internalist emphasizes that the meaning of second-order quantifiers (and other terms of the language) can be explained instead by their inferential behaviour. Thus, they propose that second-order theories can be worked with *directly*, in the object language, using one of the many available natural proof systems. Since proofs in these systems are finite and algorithmically checkable, they are perfectly respectable from a moderate perspective. By rejecting the need to interpret second-order quantifiers, internalism thus avoids one of the main problems above for modelism: how to single out full rather than Henkin interpretations.

The second characteristic idea is that informal talk of structures should be regimented, not in model-theoretic terms, but internally, within the (second-order) object language itself. Take the informal claim that there exists a natural number structure. Modelists, as before, would regiment it as a claim about models, that is, sets of a certain kind inhabiting the first-order domain, roughly as follows:

$$\exists x \ (\text{Model}\,(x) \wedge x \vDash \text{PA}_2)$$

where $\vDash$ expresses truth-in-a-model, definable in set theory. Internalists instead understand 'there exists a natural number structure', not as involving first-order quantification over models construed as objects, but rather using second-order resources:

$$\exists NzS\, \text{PA}_2\, (N, z, S)$$

where $\text{PA}_2\,(N, z, S)$ is the characteristic axiom of second-order arithmetic, PA_2, in which N, z and S have been replaced with free variables N, z and S of the relevant kind. More generally, structure-talk is not to be regimented in terms of model-theoretic interpretations; rather, it is understood in second-order terms:

> *The internalist manifesto*. For philosophical purposes, the metamathematics of second-order theories should not involve semantic ascent. Instead, it should be undertaken within the logical framework of [the] very theories under investigation. Our slogan is: metamathematics without semantics! (Button and Walsh 2018, p. 227)

How does this all bear on Putnam's Challenge? To provide a satisfying response to uniqueness, internalists must explain how, for example, PA_2 suffices to pick out a unique natural number structure, understood in terms of their proposed regimentation of structure-talk. Similarly, for determinacy, it must be explained how arithmetical claims are determined by, for instance, PA_2, using exclusively resources available to the internalist. Dedekind Categoricity and Intolerance are no longer available, as they only concern structures understood model-theoretically.

The crucial move in Button and Walsh's response to Putnam's Challenge involves an *internal* categoricity result.[13] Suppose that N_1, z_1, S_1 is an 'internal PA_2-structure', that is, $PA_2\,(N_1, z_1, S_1)$, and similarly N_2, z_2, S_2. Then the following expresses that R is an isomorphism between the two:

$$\forall x\ (N_1 x \to \exists! y\ (N_2 y \wedge Rxy))\ \wedge \forall y\ (N_2 y \to \exists! x\ (N_1 x \wedge Rxy))\ \wedge$$
$$Rz_1 z_2 \wedge \forall x \forall y\ (N_1 x \wedge N_2 y \wedge Rxy \to R\,(S_1 x, S_2 y))$$

Let us abbreviate this claim by $N_1 z_1 S_1 \overset{R}{\approx} N_2 z_2 S_2$. Note that, strictly speaking, these 'internal structures' are not objects in their own right, bearing the relation of isomorphism to each other; rather, the claim that two structures are isomorphic is shorthand for the complex higher-order claim above.

Theorem 3 (Internal Categoricity). The following is provable in second-order logic:

$$\forall N_1 z_1 S_1\, \forall N_2 z_2 S_2\ \left(PA_2(N_1, z_1, S_1) \wedge PA_2(N_2, z_2, S_2) \to \right.$$
$$\left. \exists (N_1 z_1 S_1 \overset{R}{\approx} N_2 z_2 S_2)\right)$$

Internal Categoricity and its consequences are central to the internalist response to Putnam's Challenge. The relevance to uniqueness should be clear: for if structure-talk is regimented as the internalist recommends, then Internal Categoricity says that all PA_2-structures are isomorphic. What is more, Internal Categoricity is *provable* in second-order logic alone. Unlike Dedekind's Categoricity Theorem, it does not require any set-theoretic metalanguage. Thus, internalists argue, the uniqueness of the natural number structure can be established internally, via proof, without appealing to any resources beyond second-order logic itself.

Turning from uniqueness to determinacy, Button and Walsh appeal to the following immediate consequence of Internal Categoricity, in effect an internal version of Dedekind Intolerance:

Corollary 4 (Internal Intolerance). For each formula $\phi\,(N, z, S)$ in which all quantifiers are restricted to N and all free variables displayed, the following 'Intolerance Schema' is provable in second-order logic:

[13] For Internal Categoricity, as distinct from Dedekind Categoricity, see Väänänen and Wang (2015).

(IS) $\forall NzS\ (\mathrm{PA}_2(N, z, S) \to \phi(N, z, S))$
$\vee\ \forall NzS\ (\mathrm{PA}_2(N, z, S) \to \neg\phi(N, z, S))$

Button and Walsh's core contention is that Internal Intolerance can be used by the internalist to establish the determinacy of arithmetical claims. This argument will occupy us for much of the rest of the paper, so it is worth stating in detail.

The following operator is definable in pure second-order logic:

$$\mathrm{t}\phi := \forall NzS\ (\mathrm{PA}_2(N, z, S) \to \phi(N, z, S))$$

The crucial move made by internalists is to understand this operator as expressing *determinate truth*. The idea is deceptively simple. First, on internalist grounds, $\mathrm{t}\phi$ should be roughly glossed as saying that $\phi(N, z, S)$ holds in all PA_2-structures. Second, a supervaluationist conception of determinacy is offered: ϕ is determinately true if and only if $\phi(N, z, S)$ holds in all internal PA_2-structures and determinately false if $\neg\phi(N, z, S)$ holds in all these structures.

Following Button and Walsh, let us use 'truth-internalism' to refer to this view. Truth-internalism appears to be a plausible account of determinate truth. And if it is granted, then the way appears clear for an appealing response to Putnam's Challenge: Internal Intolerance can be interpreted as saying that, for each sentence of arithmetic, it is provable in second-order logic that it is either determinately true or determinately false. As with Internal Categoricity, second-order logic is used internally, to prove claims directly, without recourse to set theory, models, or ascent to any kind of metalanguage. And since proofs in second-order logic are, by all accounts, acceptable to moderates, it looks as if internalism has done the impossible: to provide, in the face of the Löwenheim-Skolem theorems and Gödelian incompleteness, a naturalistic explanation of the uniqueness and determinacy of arithmetical structure and discourse.

In short, the internalist explanation of determinacy is as follows:

(a) For every arithmetical claim ϕ, either $\mathrm{t}\phi$ or $\mathrm{t}\neg\phi$ (from Internal Intolerance).
(b) If $\mathrm{t}\phi$, then it is determinately true that ϕ, and if $\mathrm{t}\neg\phi$, then it is determinately false that ϕ (from truth-internalism).
(c) Therefore, every arithmetical claim is either determinately true or determinately false.

In the rest of the paper, we will argue that this argument is unsuccessful: the limitative results cannot be so simply sidestepped by 'going internal'. Subsequent sections attack each premiss above. In §4 we argue that the internalist is not entitled to appeal to Internal Intolerance in the way required by (a), owing to its metalinguistic character. Internalists *are* entitled to the instances of (IS), as they are provable in PA_2. But although this might show that *each particular* arithmetical claim is determinate, it does not allow the conclusion that *every* arithmetical claim is determinate. In §5 we challenge (b) by questioning truth-internalism, and suggest that the determinate truth of a sentence of arithmetic *cannot* be understood as truth in all arithmetical internal structures, as internalists suggest. The unifying thread of our criticisms is something like this: even when moving to second-order (or more generally higher-order) theories, the Löwenheim-Skolem theorems and Gödelian incompleteness phenomena are bumps under the rug that do not go away.

4. Expressing intolerance internally

How can the internalist appeal to Internal Intolerance to establish premiss (a)? The situation might seem promising. Internal Intolerance says that, for *every* sentence ϕ in the language of PA_2, it is provable in second-order logic — and thus in PA_2 — that either ϕ or its negation holds in every internal PA_2-structure. Since our calculus is sound, it follows that *every* sentence of the language is such that either it or its negation holds in every internal PA_2-structure — that is, premiss (a) obtains. So Internal Intolerance seems to say exactly what the internalist needs.

The problem, however, is that Internal Intolerance is a *metalinguistic* result (not a theorem of second-order logic or PA_2 but a claim that infinitely many sentences of a certain form are provable in it), and thus not available to the internalist. Recall the internalist's manifesto (§3): an explanation of the determinacy of PA_2 should appeal only to what is provable in PA_2.

So Internal Intolerance shows that the internalist can assert every *instance* of (IS). And granted premiss (b), it follows that for *each* sentence ϕ, internalists can assert that it is determinately true or determinately false that ϕ.[14] But there is a big difference between this and the claim that internalists can show that, for every sentence ϕ, determinately ϕ. The difference lies in the scope of what internalists are in a position to assert: Internal Intolerance allows them to assert the determinacy of each arithmetical claim, but not that every arithmetical claim is determinate.

Although the difference may seem minor, it is highly philosophically significant. To see why, consider the analogous case of

[14] Again, we dispute the premiss in §5.

consistency statements. Let $\mathrm{Proof}_{\mathrm{PA}}$ be the usual proof predicate for first-order Peano arithmetic, PA. Compare the generalization $\forall x \neg \mathrm{Proof}_{\mathrm{PA}}(x, \ulcorner 0 \neq 0 \urcorner)$ — Con(PA), for short — to the collection of its instances, $\neg \mathrm{Proof}_{\mathrm{PA}}(\mathrm{t}, \ulcorner 0 \neq 0 \urcorner)$. The universal claim, but not the totality of the instances, is typically regarded as expressing that PA is consistent. Indeed, PA proves each instance, but by Gödel's second incompleteness theorem, if consistent, it does not prove the generalization Con(PA), a fact typically interpreted the inability of PA to prove its own consistency.

Returning to internalism: internalists wish to establish that arithmetic is determinate — that is, that *all* arithmetical sentences are either determinately true or determinately false, but Internal Intolerance entitles them at best to say, of each arithmetical sentence, that it is determinate. This is not nothing, but the generalization is still philosophically important. Indeed the internalist's position with respect to determinacy is analogous to the position of a PA-theorist with respect to consistency. The PA-theorist can establish that *this* number does not code up a proof of inconsistency, likewise *this* number, and so on …, but is limited to these piecemeal individual claims — the generalization that *no* number codes up a proof of inconsistency lies out of reach. Similarly, the internalist has the piecemeal individual claims that *this* sentence is determinate, likewise *this* sentence, and so on …, but the generalization that *no* sentence is indeterminate, corresponding to their philosophical claim about determinacy, is out of reach.

Button and Walsh (2018, §10.8) and Button (2022) recognize that this is a problem, and seek instead a substitute for Internal Intolerance that generalizes (IS) in the object language.[15] In particular, they recommend that internalists appeal to a single, *arithmetized* version of Internal Intolerance, in effect formalizing the following principle, which we'll call 'Arithmetized Internal Intolerance':

(AII) For any code of a formula $\phi(N, z, S)$ whose quantifiers are restricted to N and whose free variables are displayed, there is a code of a deduction of (IS) in second-order logic.

As it turns out, (AII) is not only expressible but also provable in PA_2. Button and Walsh conclude that internalists are able to *express* the

[15] This is clear not only from the fact that they put forward (AII) (see below) for this purpose in §10.8 of their book but from their efforts in chapter 12, where they worry about the ability of (AII) to serve this purpose, highlight the need of more suitable principles, and consider alternative ways of generalizing (IS).

determinacy of arithmetic and *establish* it deductively in PA_2, without metalinguistic ascent.[16]

Unfortunately, matters are not so simple. We argued above (and take Button and Walsh to agree) that in order to substantiate their philosophical claim that arithmetic is determinate, internalists must do more than merely establish the instances of (IS): they must establish a suitable generalization of it. This raises the question of when a sentence counts as a generalization over a class of sentences. A natural answer: when it is obtained from the members of the class using a *device of generalization*. Of course, this simply moves the question to saying what constitutes a device of generalization. This is a complex issue which is beyond the scope of this paper to answer fully, but paradigm devices include quantifiers (first- or higher-order) and certain predicates of sentences — most notably a truth predicate.[17] So, for example, the instances 'Socrates is mortal', 'Aristotle is mortal', and so on, have as a suitable generalization 'all humans are mortal'; the instances of the induction scheme $(\phi(0) \land \forall x\,(\phi(x) \to \phi(Sx))) \to \forall x\,\phi(x)$ have as a suitable generalization the second-order induction axiom $\forall X\,((X0 \land \forall x\,(Xx \to XSx)) \to \forall x\,Xx)$; and perhaps the most natural generalization of the Comprehension schema $\exists X \forall x\,(Xx \leftrightarrow \phi(x))$ is to say that any instance is *true*.

A plausible *necessary* condition on generalizations is that they entail their instances. For example, universally quantified statements entail their instances by the usual logical inference rules. Truth generalizations, on the other hand, typically entail their instances by elimination principles such as

(T-Out) $\quad \mathrm{True}(\ulcorner \phi \urcorner) \to \phi$

In claiming that (AII) generalizes the Intolerance Schema, internalists in effect attempt to use *a provability predicate* as a device of generalization — provability in second-order logic, to be more specific. But, as we now argue, there are deep reasons — related to Gödelian incompleteness — why *internalists* cannot argue this way. As we mentioned

[16] Strictly speaking, arithmetization does involve metalinguistic ascent. Taken at face value, the expressions of $\mathcal{L}_{PA_2}$ are about numbers. Only when we ascend to a metalanguage in which the coding is explicitly introduced can we give $\mathcal{L}_{PA_2}$ a syntactic reading. The issue can be easily sidestepped by working directly in a (second-order) syntax theory, for instance of the kind defended in Mount and Waxman (2021); see also Corcoran, Frank and Maloney (1974). Such a theory would in effect be a notational variant of PA_2, so no generality is lost by speaking of PA_2 as a syntax theory (as we will continue to do).

[17] For a discussion of the conditions a predicate must satisfy to function as a device of generalization, see Picollo and Schindler (2017, 2018).

before, we are tempted to regard Internal Intolerance as a generalization of (IS) because (i) it asserts that every instance of (IS) is provable in second-order logic, and (ii) from a metatheoretic perspective, we regard our calculus to be sound: we take provability in it to obey an elimination principle analogous to (T-Out), that is, a principle to the effect that, if ϕ is provable in the calculus, then ϕ.

But although this reasoning is cogent when applied to Internal Intolerance in the metalanguage, no analogous reasoning is available for (AII) in PA_2, as we will now show.

PA_2 contains a provability predicate for second-order logic, $\mathrm{Prov}_{\varnothing}(x)$, satisfying weak representability and Löb's derivability conditions.[18] But it is a direct consequence of Löb's theorem — which is actually equivalent to Gödel's second incompleteness theorem (see Smith 2013, p. 255) — that $\mathrm{Prov}_{\varnothing}(x)$ cannot obey the following principle in PA_2, on pain of triviality:

(Rfn) $\mathrm{Prov}_{\varnothing}(\ulcorner\phi\urcorner) \rightarrow \phi$

Observation 5. If PA_2 is consistent, then it does not prove all instances of (Rfn) for every sentence of the pure language of second-order logic.[19]

Proof. Note first that, if PA_2 proved all instances of (Rfn) for such sentences, it would also prove

$$\mathrm{Prov}_{\varnothing}\left(\ulcorner\mathrm{PA}_2^{\mathrm{R}}\urcorner \dot{\rightarrow} \ulcorner\phi\urcorner\right) \rightarrow \left(\mathrm{PA}_2^{\mathrm{R}} \rightarrow \phi\right)$$

where $\mathrm{PA}_2^{\mathrm{R}}$ is the ramsification of PA_2 — that is, $\exists NzS\, \mathrm{PA}_2(N, z, S)$ — and $\dot{\rightarrow}$ maps any two formulas to the conditional that has the first formula as its antecedent and the second one as its consequent. Note also that the following holds in PA_2 for each sentence ϕ of the pure language of second-order logic:

$$\mathrm{Prov}_{\varnothing}\left(\ulcorner\mathrm{PA}_2^{\mathrm{R}}\urcorner \dot{\rightarrow} \ulcorner\phi\urcorner\right) \leftrightarrow \mathrm{Prov}_{\varnothing}\left(\ulcorner\mathrm{PA}_2^{\mathrm{R}}\urcorner \dot{\rightarrow} \ulcorner\phi\urcorner\right)$$

[18] For philosophical arguments that these conditions are necessary for a provability predicate, see Halbach and Visser (2014). Where ϕ and ψ are sentences of $\mathcal{L}_{\mathrm{PA}_2}$, the Löb conditions are:

(1) If $\vdash \phi$, then $\mathrm{PA}_2 \vdash \mathrm{Prov}_{\varnothing}(\ulcorner\phi\urcorner)$

(2) $\mathrm{PA}_2 \vdash \mathrm{Prov}_{\varnothing}(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow (\mathrm{Prov}_{\varnothing}(\ulcorner\phi\urcorner) \rightarrow \mathrm{Prov}_{\varnothing}(\ulcorner\psi\urcorner))$

(3) $\mathrm{PA}_2 \vdash \mathrm{Prov}_{\varnothing}(\ulcorner\phi\urcorner) \rightarrow \mathrm{Prov}_{\varnothing}(\ulcorner\mathrm{PA}_2 \rightarrow \mathrm{Prov}_{\varnothing}((\ulcorner\phi\urcorner)\urcorner)$

Weak representability adds to (1) that if $\mathrm{PA}_2 \vdash \mathrm{Prov}_{\varnothing}(\ulcorner\phi\urcorner)$, then $\vdash \phi$.

[19] That is, where no subject-specific terms occur. Note that all instances of (IS) are themselves sentences of the pure language of second-order logic.

Thus, since PA_2 entails PA_2^R, we would have that

$$PA_2 \vdash \text{Prov}_{\varnothing}(\ulcorner PA_2 \urcorner \dot{\rightarrow} \ulcorner \phi \urcorner) \rightarrow \phi$$

However, note that $\text{Prov}_{\varnothing}(\ulcorner PA_2 \urcorner \dot{\rightarrow} x)$ expresses provability-in-PA_2, in the sense that it satisfies the relevant weak representability and Löb's derivability conditions. So by Löb's theorem, if $\text{Prov}_{\varnothing}(\ulcorner PA_2 \urcorner \dot{\rightarrow} \ulcorner \phi \urcorner) \rightarrow \phi$ is provable in PA_2, then so is ϕ. So, if PA_2 proved all instances of (Rfn), then it would have every sentence of the pure language of second-order logic as a theorem, contradicting its supposed consistency.

The point of all this is that, as a consequence of Löb's theorem (itself equivalent to Gödel's second incompleteness theorem), internalists cannot regard provability in second-order logic as a device of generalization in PA_2: in the absence of an elimination principle such as (Rfn), generalizations formulated in terms of provability — unlike truth-generalizations — do not typically entail their instances.[20] So (AII) is not a generalization of (IS) in PA_2.

A natural reply would be to move to a stronger theory in which all instances of (Rfn) *can* be proven. Such theories exist, for instance, PA_2 + (Rfn). However, generalizing the proof of Observation 5 shows that the resulting theory, if consistent, cannot be finitely axiomatizable. This is problematic for the internalist's purposes. The internalist's explanation of how a theory could be unique and determinate appeals to Internal Categoricity and Intolerance results. But these results cannot even be stated, let alone proven, for non-finitely-axiomatizable theories. So while any extension of PA_2 that yields the instances of reflection may be able to prove a generalization to the effect that PA_2 is determinate, it will be impossible for the internalist to prove the determinacy of the stronger theory in which they are working. But if the internalist wishes to answer Putnam's Challenge, they must be in a position to account for the uniqueness and determinacy of this stronger theory too, as the

[20] One possible internalist response is to argue that it suffices to find a provability predicate that sustains (Rfn), not in general, but for all *relevant* instances. In other words, since the internalist needs only to generalize over all and only the instances of (IS), is it enough to note that $\text{Prov}_{\varnothing}(x)$ sustains (Rfn) for such claims? We are sceptical. $\text{Prov}_{\varnothing}(x)$ sustains (Rfn) for this class only vacuously: since each instance of (IS) is provable in second-order logic, the relevant instance of (Rfn) is also derivable. But the fact that a predicate satisfies (Rfn) *restrictedly* is not enough for it to serve as a device of generalization. Consider $x = x$, which satisfies the Löb conditions, weak representability, and (Rfn) over the relevant class of sentences. One would hardly accept that the claim that every instance of (IS) is self-identical is a generalization of (IS). In order for a predicate to genuinely express a generalization, we need *antecedent* reason to believe that it sustains (Rfn) for the relevant class of instances.

challenge is to explain uniqueness and determinacy to the extent that they arise in mathematics in general, not just in PA_2.[21]

This section has argued that, for reasons that trace back to Gödelian incompleteness, internalists are not entitled to a suitable generalization of the Intolerance Schema. Still, as we mentioned above, they are entitled to the instances of (IS), which is not nothing. Assuming that truth in all internal PA_2-structures can be interpreted as determinate truth, these instances are enough for the internalist to establish the determinacy of any given arithmetical claim, even if the generalization over *all* such claims is out of reach. But in the next section we argue that even this partial vindication of determinacy cannot be established, because the interpretation of determinacy as truth in all internal structures cannot be sustained.

5. Explaining determinacy

For internalists, the importance of proving an internal intolerance result lies in its supposed connection with determinacy. That connection is given by what we have been calling 'truth-internalism', the claim that it is determinately true that ϕ if and only if:[22]

[21] An alternative way of generalizing over the instances of (IS) is to enrich PA_2 with a truth predicate. However, for reasons related to Tarski's indefinability theorem, this move would also require infinitely many axioms (cf. Button and Walsh 2018, §12.4), giving rise to analogous expressibility issues. Yet another way of generalizing (IS) is to move to a fifth-order language, using higher-order resources to 'code up' the notion of an interpretation of a language, allowing the definition of a *truth-in-a-structure* relation and a *truth-in-all-structures* predicate. (Moving up to an eighth-order language in turn allows similar definitions to be provided for the fifth-order language, and so on.) Button and Walsh (2018, ch. 12) consider a version of this view which, they maintain, is in the spirit of internalism. We think it's not obvious whether such a view is compatible with internalist motivations (in particular, the eschewal of semantic ascent), but we don't propose to address the issue in detail here. At any rate, this higher-order view seems equally susceptible to the objections raised in the next section. In particular, since this higher-order view is intended to embrace moderation, it faces an explanatory challenge precisely parallel to the one we press in §5.

[22] Although Button and Walsh officially gloss their operator t in terms of *truth*, they often speak interchangeably of *determinate* truth. We take it that the issue here is simply a verbal one. We can make sense of a disquotational or 'weak' notion of truth, according to which 'it is true that *p*' is equivalent to '*p*', and a 'strong' notion according to which 'it is true that *p*' is equivalent to 'it is determinately the case that *p*'. Clearly nothing turns on which notion we use; and equally clearly, the sense intended in the gloss of Intolerance is the strong notion. We will speak in terms of determinacy—indeed, Button (2022) himself speaks this way when offering a parallel presentation of internalism—but we take it that everything we say would apply equally well if reformulated in terms of truth in the strong sense.

$$\mathrm{t}\phi := \forall NzS\ (\mathrm{PA}_2\,(N,z,S) \rightarrow \phi\,(N,z,S))$$

In this section, we will argue against truth-internalism. Even waiving the discussion of the previous section, internalists cannot explain the determinacy of mathematics in the way that Button and Walsh claim. To develop the objection, however, we first need to say a little more about truth and determinacy.

5.1 Determinacy and metasemantics

The notions of determinacy and indeterminacy that arise in Putnam's Challenge and related areas of philosophy are, fundamentally, *metasemantic* or *metaconceptual* notions, concerning the content of our mathematical talk and thought.[23] For instance, Putnam speaks of what is 'fixed' by 'the total use of the language (operational plus theoretical constraints)' (1980, p. 466); similarly, for Field, ϕ is indeterminate if there is 'nothing in our inferential practice that could determine' ϕ (1998, p. 296).

Indeterminacy in this sense may have epistemic implications — if the twin prime conjecture is indeterminate, that plausibly precludes us from knowing whether it is true or false; and, to the extent we believe it's indeterminate, it is presumably pointless or irrational to continue to inquire into its truth-value. But it is not *merely* an epistemic notion. To say that a claim is determinate is to say that its truth-value is settled by the content-determining facts. So determinacy is closely tied to *metasemantic* questions, that is, about what explains or makes it the case that our words have the semantic features — meaning, content, and so on — that they do.

We believe that the metasemantic nature of determinacy motivates the following constraint (cf. Warren and Waxman 2020):

> **Metasemantic Explanatory Challenge.** Determinacy, to the extent that it arises, must be explicable.

The idea is that metasemantic properties, like determinacy (or, for that matter, indeterminacy), call out for explanation. It would be philosophically unappealing to acquiesce in an account on which facts about determinacy are brute or inexplicable. Just as it would be an abandonment of serious metasemantic theorizing to claim that words have their meaning as a matter of brute fact (not explained by, say, the conventions governing their use, or causal chains connecting them with objects in the

[23] The discussion here is couched in linguistic (as opposed to conceptual) terms simply for brevity. What we say for 'metasemantic' goes equally well for 'metaconceptual'.

world, or some such account), so too for determinacy claims. This is a domain where explanation is needed. We think that Button and Walsh's moderates will feel the force of this challenge. As Button (2022, p. 161) says, it would be 'patently ridiculous' to treat the idea that our mathematical practice pins down a unique isomorphism type as an 'inexplicable … brute feature of the world'. We are simply making the same point about determinacy.

Two clarifications are in order. First, to say that determinacy must be explicable is not to say that we must be able to actually cite an explanation of a sort that would provide illumination or understanding. We intend something more objective and idealized, perhaps: capable of explanation in principle. Readers who would prefer to put the matter in more metaphysical terms, perhaps involving grounding, are free to substitute such terminology.

Second, nothing limits the challenge specifically to mathematics. Its spirit is fully general: wherever determinacy arises, it must be explicable. For instance, someone who holds that all *physical* claims are determinate faces a similar challenge — perhaps one that can be answered in terms of conventions, causal contact, or similar. The point is that this is not an ad hoc demand that we are wheeling out to embarrass certain views of mathematics, but an essential part of a satisfying overall metasemantic picture.

5.2 Internalism cannot explain determinacy

Our objection to moderate internalism is that it is fundamentally incapable of answering this metasemantic challenge. Consider Button's succinct encapsulation:

> I affirm [PA_2] unrestrictedly and unreservedly. With Dummett, I agree that the number concept is given to us primarily in terms of *proof*. Unlike Dummett, though, I rely on an *algorithmically-checkable proof system*. Then, with the modelist, I aim to *prove* the precision of my number concept, by proving the categoricity of arithmetic. But, unlike the modelist, I am successful, and this is because my categoricity result is *internal*. (Button 2022, p. 171, emphasis in original)

But if the concept of natural number is given to us in terms of proof in some suitable recursive theory (for example, PA_2) formulated in a recursive calculus, then there will be a sentence R (for example, the Rosser sentence for PA_2) such that both R and $\neg R$ are independent of the axioms (which we assume to be consistent).[24] The internalist proves the relevant instance of intolerance, $\mathrm{t}R \vee \mathrm{t}\neg R$, which they interpret as saying that it is either determinately true or determinately false that *R*. The problem is that this interpretation leads to explanatory commitments that cannot be discharged.

From Internal Intolerance and their interpretation of t, the internalist is committed to:

(i) It is determinately true that *R* or it is determinately false that *R*.

In §5.1, we argued that determinacy claims, if true, require explanation. That is:

(ii) If it is determinately true (determinately false) that *R*, then the fact that it is determinately true (determinately false) that *R* has an in-principle explanation.

But since neither *R* nor $\neg R$ can be proven (from the internalist's recursively axiomatized theory, with a recursive notion of proof), and since these exhaust the metasemantic resources available to the internalist, we have:

(iii) Neither the determinate truth nor the determinate falsity of *R* has an in-principle explanation.

Since (i)–(iii) forms an inconsistent triad, the moderate modelist has a serious problem. In short, t cannot be interpreted as expressing determinate truth, on pain of accepting inexplicable, free-floating facts about the determinacy of our vocabulary. To see the argument's full force, it is helpful to consider some possible responses.

First, an immediate worry is that our argument overgeneralizes. It is common in mathematics to establish a disjunction without having any explanation for either disjunct. Simply consider instances of the law of excluded middle. But there is nothing necessarily problematic about such a situation. It is true that (i)–(iii) above are equally inconsistent when 'it is determinately true that *R* or it is determinately false that *R*' is replaced with an arbitrary disjunction $P \vee Q$. But in ordinary mathematical cases where a disjunction $P \vee Q$ is proven, there is no real pressure to accept the analogues of both (ii) and (iii). Which exactly should be rejected will presumably depend on the details of the case

[24] See Smith (2013, ch. 25) for details.

and one's background views of mathematics and mathematical explanations — which claims one is willing to regard as explanatorily basic, which resources one takes to be admissible in in-principle explanations, and so on. But the disjunction we think problematic for the internalist is not an ordinary mathematical case: we rely on the fact that the disjuncts are claims about *determinacy*, because such claims can't reasonably be taken as explanatorily basic — nor explained, given internalist resources.

Second, internalists might challenge (iii) by attempting to recast the point in epistemic terms. Consider an analogy from the vagueness literature. Epistemicists about vagueness hold that it is fundamentally an epistemic matter (cf. Williamson 1994). To the extent that they accept a notion of semantic determinacy, they therefore hold that all claims, even vague ones, have determinate truth-values, even if we don't (indeed possibly can't) know what they are.[25] In other words, epistemicists occupy precisely the position that this response is considering, with respect to vague claims as opposed to mathematically undecidable ones. Nevertheless, epistemicists do *not* attempt to say that the determinate truth or falsity of a particular vague claim is an inexplicable fact; rather, they recognize that there must be a metasemantic explanation of whichever disjunct obtains, at least in principle, though they take seriously the possibility that determinacy might supervene on usage in a way that is unknowable and recalcitrant to theorizing. Can internalists make a similar move, appealing to the vagaries of metasemantics to claim that the determinacy of R is indeed explained by more basic usage facts somehow we know not how?

This seems unpromising. For one thing, the view inherits one of the most troubling tensions of epistemicism: the suggestion that while nothing systematic can be known about semantic determination relations, we are nevertheless able to assert the general principle that whole classes of claims are determinate.[26] For another, going this way seems to involve abandoning the explanatory ambitions of Intolerance: the theorem would no longer be what explains the determinacy of arithmetic, as opposed to the (mysterious, unarticulated) metasemantic relations by which meaning supervenes on use. But most problematically, the internalist is if anything in *worse* shape than the epistemicist about vagueness. There is a plausible case that vague terms like 'bald' are semantically plastic, in the sense that small changes in the underlying patterns of use

[25] Many epistemicists talk as if they accept only an epistemic notion of determinacy, but this is primarily because they think that semantic determinacy is instantiated always and everywhere.

[26] See, for instance, McGee and McLaughlin (2004) for a criticism of this kind.

might lead to differences in their meaning or extension. This helps to soften the blow of the claim that the relevant semantic determination relations are inscrutable. But nothing plausibly analogous is available in the case of mathematical vocabulary: there are not *slightly* different arithmetical practices that might have the relevant kind of impact on meaning; or at least, if there are, we cannot see them.

A final rejoinder by the internalist might attempt to explain the determinacy of a given undecidable claim like *R* using *other* resources — perhaps a theory of truth, or higher-order logic — going beyond the theory — for instance, PA_2 — that they initially accepted. It should be clear by now why this would be unsatisfactory. For one thing, this move concedes that the initial theory does not suffice to establish the determinacy of arithmetic, and therefore undermines the use of Internal Intolerance. For another, the threat of regress looms as before; the determinacy of these additional resources *themselves* needs to be explained somehow. But, worst, even if this response can be made with respect to a given sentence that is undecidable in, say, PA_2, it simply cannot establish the determinacy of the whole of arithmetic. As long as the additional resources are stated in a recursive theory, and the relevant notion of proof is also recursive, then there will be sentences such that neither they nor their negations are provable. Again we are in a familiar space of alternatives: internalists are forced either to give up moderation (that is, deny that our mathematical commitments can be captured by a recursive theory or recursive notion of proof), to accept inexplicable determinacy facts, or — what we suggest — to retract the suggestion that t really does express determinate truth.

6. Uniqueness and determinacy revisited

In light of our discussion, how should we understand the philosophical significance of Internal Categoricity and Intolerance results? Although we have criticized the internalist's attempt to use Internal Intolerance to establish the determinacy of arithmetic, we have said very little to impugn either (i) the central internalist idea that informal mathematical structure-talk can be understood in second-order terms, or (ii) the use of Internal Categoricity to establish the uniqueness of the natural number structure.

We have no issue with (i) in this paper; (ii), on the other hand, although correct in a sense, requires some qualification. Granting (i), the internalist can appeal to Internal Categoricity to conclude that there are no two non-isomorphic arithmetical structures. In this sense, the internalist is entitled to conclude that PA_2 'picks out' a unique structure.

In fact, we are sympathetic to Button and Walsh's suggestion that the distinction between theories that 'pick out' *algebraic* structures (like *groups*, *fields*, *rings*, and so on) and those that pick out *univocal* structures can be elaborated as the internalist suggests, in terms of the presence of categoricity theorems.

This might be slightly puzzling. If we have established the uniqueness of some structure, should we not expect claims about it to have determinate truth-values? Not really. Although we can show in PA_2 that there is a unique arithmetical structure, our theory (indeed, *any* recursively axiomatized theory with a recursive calculus) doesn't tell us what exactly this structure looks like. In particular, for independent sentences such as the Rosser sentence R, it tells us *that* R is true in this structure or $\neg R$ is, but not *which*.

The situation can be further clarified by considering a natural account of determinacy for internalists: a claim is determinately true if it can be proven within one's theory, determinately false if its negation can be proven, and indeterminate otherwise. (Presumably everything that is provable or refutable is determinate; and presumably, in light of the explanatory constraints of §5.1, nothing else can be determinate). The internalist can prove both Internal Categoricity and the instance of Internal Intolerance for R, and can therefore accept these statements as determinately true. Internalists can, furthermore, translate between these formal results and their preferred philosophical glosses involving structure-talk, so they can claim that it is *determinately the case* that there is a unique arithmetical structure, and that R either holds in it or $\neg R$ holds in it. The point, however, is that since R is undecidable within the internalist's theory, this does *not* entail that it is determinately true or determinately false that R.[27] In other words, it is *not* possible to infer from 'it is determinately the case that R is either true in all structures or false in all structures' to 'either it is determinately the case that R is true in all structures or it is determinately the case that R is false in all structures'.

Failures of determinacy to distribute across disjunction are familiar from vagueness. In more-or-less classical theories of vagueness (for instance, supervaluationism), the existence of (unique!) sharp cut-offs in sorites sequences can be *derived*, and so on such views it is determinate that there is some unique n that is the location of the sharp cut-off. But it does not, of course, follow that for some n, it is determinate that n is the location of the sharp cut-off.

[27] Internalists cannot, of course, accept that any sentence is indeterminate in this sense, on pain of Gödelian inconsistency; but we are envisaging this argument, not as one that is run by internalists themselves, but from an external perspective attempting to get clear on the limits of the view.

We suggest that the internalist is in more or less the same position with respect to determinacy of truth-value. This is not nothing: the internalist really has shown, by their own lights, that there (determinately!) cannot be two different arithmetical structures that disagree over some sentence. But this is very far from the idea that every sentence has a determinate truth-value. In short, internalists can make a strong case that mathematics often deals with *unique* structures; but for all they have said, there is no reason whatsoever to think that *every claim about these structures* is determinate.[28]

References

Benacerraf, Paul 1973: 'Mathematical Truth'. *Journal of Philosophy*, 70(19), pp. 661–79.

Button, Tim 2022: 'Mathematical Internal Realism'. In James Contant and Sanjit Chakraborty (eds.), *Engaging Putnam*, pp. 157–82. Berlin: De Gruyter.

—— and Sean Walsh 2016: 'Structure and Categoricity: Determinacy of Reference and Truth Value in the Philosophy of Mathematics'. *Philosophia Mathematica*, 24(3), pp. 283–307.

———— 2018: *Philosophy and Model Theory*. Oxford: Oxford University Press.

Corcoran, John, William Frank, and Michael Maloney 1974: 'String Theory'. *Journal of Symbolic Logic*, 39(4), pp. 625–37.

Dedekind, Richard [1888] 1996: 'Was sind und was sollen die Zahlen?' Braunschweig: Vieweg. Translated by W. W. Beman as 'The Nature and Meaning of Numbers' in Richard Dedekind, Essays on the Theory of Numbers. Edited by Wooster Woodruff Beman. Chicago: Open Court, 1901. Revised translation in William Ewald (ed.), *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, *Volume* II, pp. 787–832. Oxford: Oxford University Press, 1996.

Field, Hartry 1980: *Science Without Numbers: A Defence of Nominalism*. Princeton, NJ: Princeton University Press.

[28] We are grateful to Andrew Brenner, Balthasar Grabmayr, Juliette Kennedy, Dan Marshall, Chris Scambler, Jared Warren, and Matteo Zicchetti for their helpful comments, and we are especially grateful to Tim Button and Sean Walsh for their stimulating work and many insightful exchanges. Thanks also to participants at meetings at the University of Haifa, the MCMP, the Eastern APA, Vita-Salute San Raffaele University, the University of Cambridge, the UNICAMP-Konstanz Set Theory Seminar, and the NUS Work in Progress Group. Thanks finally to an Editor of *Mind* and three anonymous referees for comments which considerably improved the paper.

—— 1989: *Realism, Mathematics and Modality*. Oxford: Blackwell.
—— 1998: 'Which Undecidable Mathematical Sentences Have Determinate Truth-Values?' In H. G. Dales and G. Oliveri (eds.), *Truth in Mathematics*, pp. 291–310. Oxford: Clarendon Press.
—— 2001: *Truth and the Absence of Fact*. Oxford: Clarendon Press.
Halbach, Volker, and Albert Visser 2014: 'Self-Reference in Arithmetic I'. *Review of Symbolic Logic*, 7(4), pp. 671–91.
Kreisel, Georg 1967: 'Informal Rigour and Completeness Proofs'. In Imre Lakatos (ed.), *Problems in the Philosophy of Mathematics*, pp. 138–86. Studies in Logic and the Foundations of Mathematics, 47. Amsterdam: North-Holland.
Lucas, J. R. 1961: 'Minds, Machines, and Gödel'. *Philosophy*, 36(137), pp. 112–37.
McGee, Vann, and Brian P. McLaughlin 2004: 'Logical Commitment and Semantic Indeterminacy: A Reply to Williamson'. *Linguistics and Philosophy*, 27(1), pp. 123–36.
Mount, Beau Madison, and Daniel Waxman 2021: 'Stable and Unstable Theories of Truth and Syntax'. *Mind*, 130(518), pp. 439–73.
Parsons, Charles 2008: *Mathematical Thought and Its Objects*. Cambridge: Cambridge University Press.
Penrose, Roger 1989: *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. New York: Oxford University Press.
Picollo, Lavinia, and Thomas Schindler 2017: 'Disquotation and Infinite Conjunctions'. *Erkenntnis*, 83, pp. 899–928.
———— 2018: 'Deflationism and the Function of Truth'. *Philosophical Perspectives*, 32, pp. 326–51.
Putnam, Hilary 1980: 'Models and Reality'. *Journal of Symbolic Logic*, 45(3), pp. 464–82.
Russell, Bertrand 1935: 'The Limits of Empiricism'. *Proceedings of the Aristotelian Society*, 36, pp. 131–50.
Shapiro, Stewart 1991: *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford: Clarendon Press.
Smith, Peter 2013: *An Introduction to Gödel's Theorems*, 2nd edn. Cambridge: Cambridge University Press.
Väänänen, Jouko, and Tong Wang 2015: 'Internal Categoricity in Arithmetic and Set Theory'. *Notre Dame Journal of Formal Logic*, 56(1), pp. 121–33.
Warren, Jared 2020: *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. New York: Oxford University Press.
—— and Daniel Waxman 2020: 'A Metasemantic Challenge for Mathematical Determinacy'. *Synthese*, 197(2), pp. 477–95.

Weston, Thomas 1976: 'Kreisel, the Continuum Hypothesis and Second-Order Set Theory'. *Journal of Philosophical Logic*, 5(2), pp. 281–98.
Williamson, Timothy 1994: *Vagueness*. London: Routledge.
Zermelo, Ernst 1930: 'Über Grenzzahlen Und Mengenbereiche: Neue Untersuchungen Über die Grundlagen der Mengenlehre'. *Fundamenta Mathematicæ*, 16, pp. 29–47.

