Mathematical internal realism

This paper has been submitted for inclusion in J. Conant and S. Chakraborty (eds.), *Engaging Putnam* (Harvard University Press). This is a pre-print and may be subject to minor alterations. The authoritative version should be sought in the published collection.

Tim Button
button@cantab.net
Draft of 30 January 2018

In "Models and Reality", Putnam sketched a version of his internal realism as it might arise in the philosophy of mathematics. The sketch was tantalising, but it was only a sketch. Mathematics was not the focus of any of his later writings on internal realism, and Putnam ultimately abandoned internal realism itself. As such, I have often wondered: *What might a developed mathematical internal realism have looked like?*

I will try to answer that question here, by reflecting on a discussion between Putnam, Dummett, Parsons and McGee which spanned nearly five decades. This also builds on work I have co-authored with Walsh. For readability, I have abandoned many of the historical contours in favour of "rational reconstruction", and I have relegated most of my commentary on the origins of various ideas to the footnotes of this paper. But I should like to make it perfectly clear that, without the work of the people just mentioned, this paper could scarcely have begun.

1. Acquisition and manifestation

I want to start by considering our NATURAL NUMBER concept. For clarity: I am not interested in *specific* number concepts, like THREE or TWENTY. I am interested in the *general* NATURAL NUMBER concept, as used within serious mathematics.

We have to *acquire* our mathematical concepts. Even if we are born with the capacity to acquire mathematical concepts, we are not born with the concepts themselves. No infant has the general NUMBER concept.

Equally, we must be able to *manifest* our mathematical concepts. Whilst mathematicians may sometimes work alone, mathematical practice is fundamentally communal. Mathematicians present each other with proofs and projects.[1]

In our early steps towards acquiring the NUMBER concept, we learn how to recite sequences like "1, 2, 3, 4, 5", and learn how to use such sequences to count out small collections of objects (fingers, beads, cows, or whatever). Later, we graduate to more complicated tasks, like mastering algorithms for adding (or multiplying) numbers in decimal

[1] Manifestation and acquisition are deep themes throughout Dummett's work (e.g. 1963: 188–90).

notation. But my interest here is not in numerical cognition, infant or adult. It is in the NUMBER concept itself, as used in *serious mathematics*. And, whatever developmental-cum-pedagogical steps we must take *towards* acquiring that concept, we have acquired it only when we have grasped some full-blown mathematical *theory*, such as Peano Arithmetic.[2] Equally, we manifest our grasp of the concept by using some such theory.

So, for the rest of this paper I will assume both that *mathematical concepts can be (and only can be) fully acquired by mastering a theory*, and that *mathematical concepts can be (and can only be) fully manifested by presenting a theory*.[3]

2. The modelist answer

With all this assumed, I want to raise a question:[4]

> *How precise is our* NATURAL NUMBER *concept?*

I want to show how the relationship of theories to (the acquisition and manifestation of) concepts threatens to *constrain* the precision of our mathematical concepts. To explain how the threat arises, I will introduce a specific philosophical character, the *modelist*. Her position is extremely tempting, but it is ultimately untenable.

The modelist answers my initial question with a slogan:

> **Modelist.** The NATURAL NUMBER concept is precise up to isomorphism.

But, of course, her slogan needs to be explained:

> **Modelist.** To consider the NATURAL NUMBER concept, we can simply consider the class of all natural-number sequences. After all, that class encodes everything we could ever want to know about the NATURAL NUMBER concept. So, when you ask, "How precise is our NATURAL NUMBER concept?", I attack this by instead asking, "How refined is the class of arithmetical models?"
>
> Well, on the one hand: suppose we had two sequences that were not isomorphic. In that case, we would not allow that *both* were natural-number sequences, since they would differ in some arithmetically important respect. So: every model in the class must be isomorphic to every other.
>
> On the other hand: arithmetic does not really seem to care about the differences between isomorphic sequences. So: the class should be closed under isomorphism.
>
> Combining these two points: every model in the class must be isomorphic to every other, and the class must be closed under isomorphism. In short, the class of arithmetical models is an *isomorphism type*.[5]

---

[2] For an interesting discussion concerning the stage at which we (implicitly) come to grasp Peano Arithmetic (or something like it), see Rips et al (2008 and the subsequent 'Open Peer Commentary').

[3] In fact, I only require that this assumption holds for the concepts NUMBER and SET. Even those who are not yet convinced by this assumption may be interested to see where it leads.

[4] Dummett (1963) and Parsons (1990) ask roughly this question. Putnam (1980) raises very similar issues, but via questions which focus more on *objects* than on *concepts*. However, objectual and conceptual versions of the question are very similar (see Button & Walsh, 2018: chs.6–8); so, for simplicity, I will focus solely on the conceptual version.

That is what I mean, when I say that the NUMBER concept is precise up to isomorphism. I mean: we can (and should) use an isomorphism type as a surrogate for the NUMBER concept.

Note that many mathematical concepts are not this precise. As an example: the LINEAR ORDER concept is a perfectly decent concept, but plenty of linear orders are *not* isomorphic, so that the LINEAR ORDER concept is not precise up to isomorphism. My view is roughly that our *foundational* mathematical concepts are (or, aim to be) precise up to isomorphism. Admittedly, the idea of a "foundational" concept is a little imprecise, but I hope you get a sense of my ambition.

That is modelism, in a nutshell. Modelism is obviously structuralist. However, modelism is just one *version* of structuralism. And its special reliance on model theory gives rise to its name, *modelism*.[6]

Modelism is appealing. Unfortunately, as Putnam taught us, it succumbs to a *model-theoretic argument.*[7]

In §1, I insisted that mathematical concepts must be tied to theories, via manifestation and acquisition. So, if the NUMBER concept is precise up to isomorphism, as the modelist insists, then our arithmetical theory must pick out an isomorphism type. But formal theories are offered in formal languages, and formal languages have certain provable limitations. For example, we have:

The Löwenheim–Skolem Theorem. If a (countable, first-order) arithmetical theory has any infinite models, then it has models of every infinite cardinality.
A Corollary of Compactness. If a (first-order) arithmetical theory has any infinite models, then it has models containing non-standard elements.

So – assuming we are limited to (countable) first-order theories – our theory cannot pick out an isomorphism type. But then, given that the NUMBER concept was supposed to be precise up to isomorphism, no *theory* will allow us (fully) to manifest or acquire our NUMBER concept. And that contradicts what I insisted upon in §1.

This is the kernel of the model-theoretic argument against modelism. To make it stick, though, we must defend the assumption that the modelist is limited to considering formal, (essentially) first-order, theories.

First, then, consider *formality*. Arithmetic, as a practice, is not just a list of axioms, but rather a "MOTLEY [of techniques and proofs".[8] So, a modelist might propose that this motley plays some role in picking out an isomorphism type.[9]

---

[5] Our modelist might do better to focus on *definitional equivalence* instead of *isomorphism* (see Button & Walsh 2018: §§5.1–5.2); but this would not change the dialectic.
[6] Button & Walsh (2018: ch.6) coined the term, and go into more detail.
[7] The remainder of this section presents the central problem I extract from Putnam's (1980) invocation of the Löwenheim–Skolem Theorem. (Admittedly, Putnam raised the issue in a more "objectual" than "conceptual" key; but see footnote 1, above.) Dummett (1963: 192) raised a similar problem, focussing on Gödelian incompleteness. For more, see Button & Walsh (2018: ch.7).

Now, insofar as model theory (as a branch of pure mathematics) considers theories, it considers only *formal* theories. So, if a modelist appeals to informal mathematics, then we cannot raise problems for her *just* by employing results from model theory. And this might seem like a strike in favour of an "informalist" modelism.

However, this point really cuts both ways. The very notion of an ISOMORPHISM TYPE is something we define within model theory. So it is hard to see how anyone could even *hope* to explain how an *informal* theory could pin down an isomorphism type. Moreover, leaving this issue *unexplained* is not viable, since the following two responses seem equivalent, and hence equally absurd:

(i) It is just a brute fact – a "primitive, surd, metaphysical truth"[10] – that our informal mathematical practice pins down a particular isomorphism type.
(ii) Luckily, everyone who wears this *particular* motley just happens to pick out a very *specific* thing, even though (*a priori*) any of us might have picked out *different* things, or indeed have failed to pick out *anything at all*.

As such, I take it that modelists are restricted to using formal theories, and that they must explain how formal theories can pin down isomorphism types. [11]

The model-theoretic argument against modelism did not, though, *just* assume that the modelist's favourite theory must be formal; it also assumed that the theory must be (essentially) first-order.

A little technical background will help to explain this point. When we use the full semantics for second-order logic, we treat second-order quantifiers as ranging over the full powerset of the first-order domain. So, we gloss "$\forall X$" roughly as "for any subset of the first-order domain". Now, neither the Löwenheim–Skolem nor the Compactness theorems hold for full second-order logic. On the contrary, we have the following: [12]

Dedekind's Categoricity Theorem. Second-order Peano arithmetic is categorical in full second-order logic (i.e., all models of the theory are isomorphic).

---

[8] Wittgenstein (1956: §46); "MOTLEY" is a translation of "*buntes* GEMISCH".
[9] This seems to be Benacerraf's (1985: 108–11) response to Putnam (1980).
[10] Putnam (1981: 46)
[11] Admittedly, mathematicians were discussing "the natural numbers" long before they had any *formal* theories (in the modern sense). So, to tell the *historical* story of how we (collectively) acquired the NUMBER concept, we would certainly need to talk at length about informal practice. I do not, though, think this affects the general point that, in terms of §1, the modelist should concede that she must employ a *formal* theory if she wants to manifest the concept in its *full precision*.
[12] For a modern proof, and references to plenty of others, see e.g. Button & Walsh (2018: §7.4).

So, invoking this result, we can expect the modelist can reply to the model-theoretic argument as follows:

> **Modelist.** The theory of *second-order* Peano arithmetic allows us to acquire and manifest a NUMBER concept that is precise up to isomorphism.

This reply is tempting, but fatally flawed.[13] The flaw does not concern the use of *second-order* Peano arithmetic (there is nothing intrinsically wrong with concatenating quantifiers with upper-case letters). The flaw concerns the appeal to the *full semantics* for second-order logic.

Our modelist wants to say that some (formal) theory allows us to acquire and manifest our NUMBER concept. Indeed, she has *specified* a particular theory: second order Peano arithmetic. However, if we approach second-order Peano arithmetic using the *Henkin*-semantics for second-order logic, then both the Löwenheim–Skolem and Compactness results return. So, the modelist must insist that we approach second-order Peano arithmetic using a particular semantics; the *full* semantics.

At this point, we must ask her to explain how we acquire and manifest the concepts involved in that *semantics*. I expect this reply:[14]

> **Modelist.** The key concept, i.e. POWERSET, just comes down to the idea of ALL COMBINATORIALLY POSSIBLE SUB-COLLECTIONS OF A COLLECTION.

This is true. But we are no more born with that *general* mathematical concept, than we are born with the general NUMBER concept; we must *acquire* it. Equally, we must be able to *manifest* it. The rules of §1 apply.[15]

Now, in §1, I noted that counting out small collections of objects is probably an important step on the road towards acquiring the NUMBER concept. In the end, though, I insisted that we grasp the *general* concept only when we grasp some full-blown mathematical theory. Similarly: manipulating small collections of objects may be an important step on the road towards acquiring the notion of SET, but we grasp the *general* concept of POWERSET only when we grasp some full-blown mathematical theory.

As above: allowing this theory to be *informal* will leave everything unexplained. So the modelist must accept that the theory which gives us the POWERSET concept is *formal*.

Now, though, the modelist has begun on an infinite regress. To make it explicit:

(1a) To explain how we come to grasp the NUMBER concept, the modelist presents us with a formal theory, T_1.

---

[13] What follows is, in effect, one version of Putnam's famous just-more-theory manoeuvre; see Putnam (1977: 486–7; 1980: 477, 481) and references and discussion in Button (2013: chs.4–7) and Button & Walsh (2018: §§2.3, 7.7–8).

[14] Thanks to Mary Leng for suggesting this way of putting it.

[15] It is sometimes suggested that our grasp of *plural logic* will deliver the required combinatorial concept. But the same question arises: what allows us to grasp *full* plural logic, rather than *Henkin* plural logic? Florio and Linnebo (2016) develop this criticism elegantly.

(1b) However, if T_1 is to pin down the NUMBER concept up to isomorphism, T_1 must be understood via some "intended" semantics.

(1c) So, if T_1 is to achieve what the modelist wants, we must understand the concepts involved in T_1's "intended" semantics *before* we are introduced to T_1.

(2a) To explain how we come to grasp those semantic concepts, the modelist presents us with a formal theory, T_2.

(2b) However, if T_2 is to pin down those semantic concepts sufficiently precisely, T_2 must be understood via some "intended" semantics.

(2c) So, if T_2 is to achieve what the modelist wants, we must understand the concepts involved in T_2's "intended" semantics *before* being introduced to T_2.

…

So it goes. This is clearly a *regress*. Equally clear, it is a *vicious* regress. It simply cannot be a constraint, on acquiring or manifesting the concepts involved in one theory, that we must *first* acquire or manifest the concepts involved in the theory at the next meta-level. [16] For then we would never be able to acquire or to manifest our concepts *at all*.

One final point. Earlier, our modelist moved straight from first-order logic to second-order logic with its full semantics. In fact, she might have invoked one of several alternative logics in a similar effort to rebut the model-theoretic argument. But there is a hard limit here. As noted, we can present a model-theoretic argument using just the Compactness Theorem. But Compactness holds for any logic with a finitary (sound and complete) proof system. [17] So: if the modelist wants to use a logic which is strong enough to pin down an isomorphism type, then the logic cannot be fully articulated in terms of a proof system, but must be articulated semantically. And that suffices to set the modelist off on her vicious regress.

3. A Dummettian approach

Modelism has failed. We need an alternative. An obvious thought is to approach matters proof-theoretically, rather than model-theoretically. Indeed, this was Dummett's approach. His central thought is something like this:

(a) Mathematical concepts are determined fully by their uses in proofs.

The hope is that this will do better than modelism, in coping with the requirements of acquisition and manifestation from §1. After all, rules of proof are rather more tractable than isomorphism types, when it comes to teaching and learning mathematics.

Unfortunately, there is an immediate barrier to this proposal. Let P be any *computable* system of proof, i.e. a system for which there is an algorithm which decides whether any

[16] Note that it is useless to suggest that $T_n = T_{n+k}$, for some n and all k, since there are guaranteed to be "unintended" Henkin-style interpretations of each T_n, and these will yield unintended interpretations of T_1.

[17] See Button & Walsh (2018: §7.9).

putative proof is a genuine proof. Now, suppose (for reductio) that *P*-provability exhausts the arithmetical facts; i.e. that, for any arithmetical sentence φ:

$$\varphi \textit{ iff } \text{there is a } P\text{-proof of } \varphi$$

Since *being a P-proof* is decidable, some computable function captures the idea that *x* is (a code of) a *P*-proof of (the code of) φ. Using this, we can formulate an *arithmetical* predicate, *Tr*, such that, for any arithmetical sentence φ:

$$\text{there is a } P\text{-proof of } \varphi \text{ iff } Tr(`\varphi\text{'})$$

It follows that, for any arithmetical sentence φ:

$$\varphi \textit{ iff } Tr(`\varphi\text{'})$$

But this contradicts Tarski's Indefinability Theorem.[18] So, we must retract the assumption that *P*-provability exhausts the arithmetical facts. More generally, we must accept that

(b) No computable system of proof exhausts the arithmetical facts.

Dummett is aware of this sort of reasoning;[19] but he does not take it to undermine (a). Instead, since he insists that the NUMBER concept *is* fully determined by its use in proofs, he takes (b) to show that "no *formal* system can ever succeed in embodying all the principles of proof that we should intuitively accept".[20] That is, in accordance with (a), the NUMBER concept is fully determined by its use in intuitively acceptable proofs; but, by (b), the notion of INTUITIVELY ACCEPTABLE PROOF is *non*-computable. From this, Dummett concludes that that the NUMBER concept itself "cannot be fully expressed by means of any formal system".[21]

Unfortunately, this leads to a rather unhappy conclusion. Following Dummett, I have insisted that our NUMBER concept must be both acquirable and manifestable. But *machines*, I take it, can only acquire concepts which *can* be fully expressed by means of some formal system.[22] Given Dummett's claim that the NUMBER concept "cannot be fully expressed by means of any formal system", he must accept that machines cannot acquire the NUMBER concept, but only acquire some imprecise approximation to it. In short, Dummett's position commits us to a startling disjunction:

[18] I have put the problem this way, rather than simply invoking the fact that *the class of arithmetical truths is not computably enumerable*, to emphasise that the problem does not depend on invoking a notion of arithmetical truth that is (somehow) "prior" to a notion of proof.
[19] Though Dummett (1963) focusses on Gödelian reasoning, rather than on Tarskian undefinability.
[20] Dummett (1963: 200).
[21] Dummett (1963: 186).
[22] Setting aside machines with access to oracles.

(c) Either we are not machines, or we do not possess the NUMBER concept.[23]

I cannot really take seriously the possibility that we do not possess the NUMBER concept. Equally, though, I do not want my philosophy of mathematics to require that we are not machines. Whilst much more could be said on both fronts, I ultimately have no option but to part ways with Dummett.

4. The Skolem–Gödel Antinomy

The previous three sections can be summarised as follows.

We acquire and manifest our mathematical concepts via formal theories. Modelism treats such theories model-theoretically, but it succumbs to a model-theoretic argument. The obvious alternative is to treat formal theories proof-theoretically. If we want to allow for the possibility that we are machines, then the relevant system of proof must be *computable*. But our NUMBER concept is sufficiently precise and detailed that no computable system of proof exhausts the arithmetical facts.

All told, we find ourselves in the following bind:

The Skolem–Gödel Antinomy. *Our mathematical concepts are perfectly precise. However, these perfectly precise mathematical concepts are manifested and acquired via a formal theory, which is understood in terms of a computable system of proof, and hence is incomplete.*

One might well worry that *something must have gone wrong*, because a NUMBER concept articulated in an *incomplete* theory *must* thereby be *imprecise*. I certainly feel the tension; indeed, that is why I call this predicament an "antinomy".[24] Still, I do not think that anything has gone wrong. This *really is* our predicament, and we need to face up to it.

With that in mind, the rest of this paper outlines a position, *internalism*, which aims to resolve the Skolem–Gödel Antinomy. Moreover, I think that internalism provides a detailed development of the mathematical internal realism which Putnam sketched towards the end of his "Models and Reality".[25]

---

[23] This is obviously similar to Gödel's Disjunction (1951: 310). However, Dummett's right disjunct ("we do not possess the NUMBER concept") is much stronger than Gödel's ("there exist absolutely unsolvable diophantine problems").

[24] Cf. Putnam's (1980: 464) use of "antinomy".

[25] I should clarify my ambitions here, not least because the material in the second half of this paper has its basis in joint work with Sean Walsh (Button & Walsh 2018). In that book, Sean and I were not *endorsing* internalism; we simply wanted to articulate the *best possible version* of internalism. In this paper, I want to stick my neck out a bit (but only a bit) further.

I am confident that the Skolem–Gödel Antinomy accurately describes our predicament. Moreover, internalism strikes me as the most promising line of response to the Antinomy. Indeed, at the moment, I see no other way to face up to the Antinomy.

5. Internalism about arithmetic

I will start by outlining a formal theory of arithmetic which articulates the NATURAL NUMBER concept *incompletely*, but still shows that concept to be perfectly *precise*.

I do not want my theory to assume that everything is a number. So I need a primitive predicate, '$N(x)$', to be read as 'x is a natural number'. I also need a primitive function symbol, '$s(x)$', to be read as 'the successor of x'. To save some space in my formalisms, I will introduce two obvious abbreviations:

$(\forall x : \Psi)\varphi$ abbreviates $\forall x(\Psi(x) \rightarrow \varphi)$
$(\exists x : \Psi)\varphi$ abbreviates $\exists x(\Psi(x) \wedge \varphi)$

And now, using these symbols and abbreviations, I can lay down four axioms:

(1) $(\forall x : N)\, N(s(x))$
i.e. the successor of any number is a number

(2) $(\exists z : N)(\forall x : N)\, s(x) \neq z$
i.e. there is a 'zero' element

(3) $(\forall x : N)(\forall y : N)(s(x) = s(y) \rightarrow x = y)$
i.e. successor is injective on the numbers

(4) $\forall F(\{(\forall z : N)[(\forall x : N)s(x) \neq z \rightarrow F(z)] \wedge (\forall x : N)[F(x) \rightarrow F(s(x))]\} \rightarrow (\forall x : N)\, F(x))$
i.e. induction. More fully: for any property F, if every 'zero' element has F and F is closed under successor, then every number has F.

Let PA_{int} be the conjunction of these four axioms. The name abbreviates Peano Arithmetic, *int*ernalized, and PA_{int} is just ordinary second-order Peano Arithmetic, with all the axioms relativized to 'N'. This is the theory which I will wield in the face of the Skolem–Gödel Antinomy.

Still, there is more work to be done to clarify internalism (on which, see §9). And, though I hope otherwise, such further work may end up exposing deep flaws in internalism.

So, the situation is this. If you *forced* me to declare for *some* position in the philosophy of mathematics, then I would declare myself an internalist, and hope that everything works out for the best. But, absent that compulsion, I hesitate to call myself an *avowed* internalist. For readability, though, I will keep these reservations buried in this footnote. In the main text of this paper, I will write as a straightforward *advocate* of internalism.

Perhaps I am being absurdly cagey. But, I am mindful of Putnam's (2000: 127–8) remark: "This identification of truth with superassertability is one that I myself found somewhat implausible, but at that time (the late 70s and early 80s) I did not see how to make sense of the notion of truth in any other way, given the failure of metaphysical realism."

To appreciate the virtues of PA_{int}, I want you to imagine two people, Solange and Tristan, who have both learned PA_{int}.[26] They are now happily babbling away to each other, exploring its consequences.

Although they will presumably use the *same* words as each other, to keep things clear, I will use 'N_1' for Solange's number-predicate and 's_1' for her successor-function, so that Solange advances PA_{int} in this subscripted vocabulary. I will call her subscripted theory $PA(N_1, s_1)$. Similarly, I will have Tristan advancing $PA(N_2, s_2)$.

In advancing $PA(N_1, s_1)$ and $PA(N_2, s_2)$, there is no guarantee that Solange and Tristan are talking about the *same objects* (if they even think of themselves as talking *about* objects at all). To take a trivial example: maybe "Solange's zero element" is Solange herself, and "Tristan's zero element" is Tristan, so that Solange can rightly say "zero is hungry", whilst Tristan rightly says "zero is not hungry". But this *is* trivial, and for an obvious reason: mathematicians basically only care about arithmetical features of the natural numbers, and not (for example) whether the numbers are *hungry*. We philosophers should probably do the same.

In that case, the *important* question is this:

Is there any guarantee that Solange's numbers and Tristan's numbers share the same arithmetical structure?

In fact, PA_{int} is just strong enough to give us this guarantee. We have:[27]

Internal Categoricity of PA.

$$\begin{aligned} \vdash \forall N_1 \forall s_1 \forall N_2 \forall s_2([PA(N_1, s_1) \land PA(N_2, s_2)] \to \\ \exists R\, [\forall v \forall y(R(v, y) \to [N_1(v) \land N_2(y)]) \land \\ (\forall v : N_1)\exists! y\, R(v, y) \land \\ (\forall y : N_2)\exists! v\, R(v, y) \land \\ \forall v \forall y(R(v, y) \leftrightarrow R(s_1(v), s_2(y)))]) \end{aligned}$$

Roughly, this says the following: given that Solange's number-property and successor-function behave PA_{int}-ishly, and so do Tristan's number-property and successor-function, there is some relation, R, which takes us from Solange's numbers to Tristan's, and is bijective, and preserves successor (and hence also preserves zero-hood). Or, more briefly:

---

[26] The idea here is inspired by Parsons's (1990; 2008) discussions of Kurt and Michael. For more on the similarities and differences between this approach and Parsons's, see Button & Walsh (2018: §10.B).

[27] See Button & Walsh (2018: §10.B) and Väänänen & Wang (2015, Theorem 1). For the sake of exposition, I have moved freely between treating e.g. "N_1" as a predicate and treating it as a relation-variables, leaving it to context to individuate what treatment is appropriate. For a rigorous treatment, see Button & Walsh (2018: chs.10–12).

Provably, all of Solange's arithmetical structure is mirrored in Tristan's numbers, and vice versa.

This *internal* categoricity result bears a family resemblance to Dedekind's categoricity result, mentioned in §2, that all models of second-order Peano arithmetic are isomorphic. But it is worth spelling out the extremely important differences between these results.

Dedekind's result is model-theoretic. It is stated and proved in a semantic metalanguage. The internal categoricity result, by contrast, amounts to *metamathematics without semantic ascent*. It involves no *semantic* considerations at all. The turnstile '$\vdash$' which I used in stating the result indicates that the proof takes place in the ordinary *deductive* system for (impredicative) second-order logic. The proved sentence is in the same language as $\mathrm{PA}_{\mathrm{int}}$ itself (or, more austerely, in the "purely logical" fragment of $\mathrm{PA}_{\mathrm{int}}$). So it is an *internal* categoricity theorem, in precisely the sense that it does not take us beyond the object language itself, or outside that object-language's deductive system.

Sticking with (mere) deduction has a *benefit*. In §2, our modelist attempted to invoke Dedekind's categoricity result. As such, she needed to invoke a semantic theory, and this set her off on a vicious regress. Since no semantic theories are involved in the Internal Categoricity Theorem, no similar regress can arise.

However, sticking with (mere) deduction also comes with a *cost*. Inevitably, $\mathrm{PA}_{\mathrm{int}}$ fails to prove its own Gödel-sentence. Since we are viewing $\mathrm{PA}_{\mathrm{int}}$ deductively, we must therefore regard it as *incomplete*.

All of this, of course, was promised us by the Skolem–Gödel Antinomy. Nonetheless – and to address that Antinomy – I now want to insist that $\mathrm{PA}_{\mathrm{int}}$ succeeds in introducing a perfectly *precise* NUMBER concept.

To explain why, I want to revisit Solange and Tristan, respectively affirming $\mathrm{PA}(N_1, s_1)$ and $\mathrm{PA}(N_2, s_2)$. Suppose that Solange affirms an appropriate formalisation of "every even number is the sum of two primes", to which Tristan shakes his head and affirms an appropriate formalisation of "some even number is not the sum of two primes". Ask yourself:

Is there any guarantee that Solange and Tristan are in genuine disagreement?

We already noted that Solange and Tristan need *not* agree about what the numbers *are*. Still, Goldbach's Conjecture is purely arithmetical. So, since all of Solange's arithmetical structure is mirrored in Tristan's numbers (and vice versa), we might expect there to be *genuine disagreement* here. And that is exactly what we find. More precisely, we have the following corollary of $\mathrm{PA}_{\mathrm{int}}$'s internal categoricity:[28]

[28] For a full statement and proof, see Button & Walsh (2018: §§10.5, 10.B). Note the *schematic* character of this result. This might lead us to ask the internalist questions about the *syntactic theory* (as we asked the modelist questions about the *semantic* theory), but I think these can be addressed (see Button & Walsh, 2018: §10.8).

Intolerance of PA. For each second-order formula φ, whose only free variables are N and s, and whose quantifiers are all restricted to N:

$$\vdash \forall N \forall s(\mathrm{PA}(N, s) \rightarrow \varphi) \vee \forall N \forall s(\mathrm{PA}(N, s) \rightarrow \neg\varphi)$$

And, if you stare at this for a minute or so, you should see that this can be informally glossed as follows:

On pain of provable inconsistency, no two PA_{int}-ish NUMBER *concepts can diverge over any arithmetical claim.*

I call this an *intolerance* theorem, since it shows that $\mathrm{PA}_{\mathrm{int}}$ does not tolerate different ways of pursuing arithmetic. If Solange affirms Goldbach's Conjecture whilst Tristan denies it (in their respective languages), then Solange cannot just shrug and say: "that might hold in *your* numbers, but it doesn't hold in mine!" If they share a logical language, then – on pain of inconsistency – they must hold that one of them is *wrong*.

This observation is the key to my claim that $\mathrm{PA}_{\mathrm{int}}$ articulates the NUMBER concept precisely. To spell out the final steps, I propose that we should think about *precision* in roughly the way that supervaluationists think about determinacy, i.e. via this heuristic:

If we can equally well render a claim right or wrong, just by sharpening up the concepts involved in the claim in different ways, then that claim is *indeterminate* (prior to any sharpening of concepts). Otherwise, it is *determinate*.

Now let φ be any arithmetical claim. If φ holds for *every* $\mathrm{PA}_{\mathrm{int}}$-ish NUMBER concept, then we cannot render φ right or wrong, just by considering *Solange's* number concept rather than *Tristan's*, or whatever. So, by the above heuristic, it is *determinate* that φ. More generally, this suggests that we should gloss $\forall N \forall s(\mathrm{PA}(N, s) \rightarrow \varphi)$ as 'it is determinate that φ'. And this allows us to restate the Intolerance Theorem as follows:

Glossed Intolerance. For each second-order formula φ, whose only free variables are N and s, and whose quantifiers are all restricted to N:

$$\vdash \forall N \forall s(\mathrm{PA}(N, s) \rightarrow \varphi) \vee \forall N \forall s(\mathrm{PA}(N, s) \rightarrow \neg\varphi)$$

i.e.: *either it is determinate that φ or it is determinate that $\neg\varphi$*

i.e.: *it is determinate whether φ*

In sum: thanks to its intolerance, $\mathrm{PA}_{\mathrm{int}}$ articulates our NATURAL NUMBER concept sufficiently precisely, that *every arithmetical claim is determinate.*

Allow me to summarise this section. The theory $\mathrm{PA}_{\mathrm{int}}$ has just four "axioms" – its conjuncts – so that there is no difficulty in *acquiring* or *manifesting* either the theory itself or the concepts it articulates. Plenty of arithmetical claims are not decided by $\mathrm{PA}_{\mathrm{int}}$; it articulates the NUMBER concept *incompletely*. But PA_{int} articulates our NUMBER concept sufficiently *precisely*, that (provably) every arithmetical claim is determinate.

In short, PA_{int} gives us a way to respond to the Skolem–Gödel Antinomy of §4, in the specific case of the NUMBER concept. That is the response I want to offer. And here is a statement of my position, more generally:

> **Internalist (about arithmetic).** I affirm PA_{int} 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 upon a *computable* system of proof. Then, with the modelist, I aim to *prove* the precision of my NUMBER concept, by proving the categoricity of my arithmetical theory. But, unlike the modelist, I am successful; and I succeed, because *my* categoricity result is *internal*.

6. Internalism about set theory

There is *much* more to say about internalism about arithmetic. I will say some of it in §9. First, I want to consider internalism about set theory. In brief, I want to lift the story of §5 over from the NUMBER concept to the SET concept.

As in the previous section, I will start by introducing an "internalized" theory of pure sets. Rather than using a Zermelo–Fraenkel-style theory, I will use a set theory which captures the "minimal core" of the cumulative iterative notion of set, namely, Scott–Potter set theory.[29] (I will explain this talk of the "minimal core" in a moment.)

Mine will be a theory of *pure* sets. So, if there is a set of the cows in the field, then I will simply *ignore* it. To restrict attention to pure sets in this way, I need a predicate, '$P(x)$', to be read as 'x is a pure set'. Unsurprisingly, I will also need a membership predicate, '$\in$'. And, using these symbols, and the abbreviations of §5, I can write down some axioms:

(1) $\forall x \forall y(x \in y \to (P(x) \land P(y)))$
i.e. we restrict our attention to membership facts between pure sets

(2) $(\forall x : P)(\forall y : P)[\forall z(z \in x \leftrightarrow z \in y) \to x = y]$
i.e. pure sets are extensional entities

(3) $\forall F((\exists x : P)\forall z(z \in x \leftrightarrow F(z)) \leftrightarrow (\exists y : Level)(\forall z : F) z \in y)$
i.e. for any property F: there is a pure set whose members are exactly the instances of F iff there is some "level" such that every instance of F is a member of that "level". Or, more punchily: a property determines a pure set iff all its instances are members of some "level".

As written, principle (3) uses an undefined expression, "*Level*". However – and this is the neat trick about the Scott–Potter approach – we can explicitly define "*Level*" in terms of set-

[29] This is the core of the set theory presented by Potter (2004, especially ch.3). For a brief presentation of all that is required for the purposes of this paper, see Button & Walsh (2018: §§8.B–C, 11.C–D).

membership.[30] As such, the only primitives we need are "P" and "$\in$". Let SP_{int}, (for *S*cott–*P*otter, *int*ernalized) be the conjunction of these three axioms.

Crucially, SP_{int} proves: the levels are well-founded by membership. Combining this with principle (3), this is why SP_{int} gives us the "minimal core" of the cumulative iterative conception of sets. It is the "core", since it tells us that sets are stratified into well-ordered levels. It is "minimal", because it makes no comment at all about how far the sequence of levels runs. (There is no powerset axiom; no axiom of infinity; no axiom of replacement.)

Indeed, thinking model-theoretically, the "pure parts" of the full second-order models of SP_{int} are (up to isomorphism) exactly the (arbitrary) stages of the cumulative hierarchy, as described by second-order Zermelo–Fraenkel set theory.[31] But I mention this fact, only to make SP_{int} feel a bit more familiar. I will treat SP_{int} deductively, as I treated PA_{int} in §5.

Working deductively, then, we can recover an "internal" counterpart of Zermelo's *quasi*-categoricity theorem. Roughly, this says: if both Solange's and Tristan's pure sets behave SP_{int}-ishly, then their sets are isomorphic *as far as they go*, but Solange's might go further than Tristan's (or vice versa). However, to keep this paper short, I will leave the details of internal quasi-categoricity for elsewhere,[32] and skip straight ahead to a set theory which is internally (totally) *categorical*. I call this theory CSP_{int}, for *C*ategorical *S*cott–*P*otter set theory. We obtain it by adding a fourth conjunct to SP_{int}:

(4) $\exists f\,(\forall x\, P(f(x)) \wedge \forall y(P(y) \rightarrow \exists! x\, f(x) = y))$
i.e. there are exactly as many pure sets as there are objects simpliciter, i.e. as objects which are either pure sets or not. (The quantifier "$\exists f$" here is second-order.)

In the deductive system for impredicative second-order logic, we can then prove internal categoricity for CSP_{int}. Informally, this is the result that there is a membership-preserving bijection from Solange's pure sets to Tristan's. Formally: [33]

Internal Categoricity of CSP.

$$\begin{aligned}
\vdash \forall P_1 \forall \in_1 \forall P_2 \forall \in_2 ([\text{CSP}(P_1, \in_1) \wedge \text{CSP}(P_2, \in_2)] \rightarrow \\
\exists R\,[\forall v \forall y (R(v,y) \rightarrow [P_1(v) \wedge P_2(y)]) \wedge \\
(\forall v : P_1) \exists! y R(v,y) \wedge \\
(\forall y : P_2) \exists! v R(v,y) \wedge \\
\forall v \forall x \forall y \forall z ([R(v,y) \wedge R(x,z)] \rightarrow [v \in_1 x \leftrightarrow y \in_2 z]])
\end{aligned}$$

[30] I omit the definition; for details, see Potter (2004: 24, 41) and Button & Walsh (2018: §§8.5, 8.B).
[31] See Button & Walsh (2018: §8.C).
[32] Interested readers should look to Button & Walsh (2018: §§11.2, 11.C).
[33] See Button & Walsh (2018: §§11.4, 11.D).

From *internal categoricity*, we can also obtain *intolerance*. Informally, this result says that, on pain of inconsistency, no two CSP_{int}-ish set concepts can diverge over any pure set-theoretic claim. Formally: [34]

> **Intolerance of CSP.** For each second-order formula φ, whose only free variables are P and $\in$, and whose quantifiers are all restricted to P:
> $$\vdash \forall P \forall \in (CSP(P, \in) \rightarrow \varphi) \vee \forall P \forall \in (CSP(P, \in) \rightarrow \neg\varphi)$$

The situation, then, is as with PA_{int}. The theory CSP_{int} gives *internalists about set theory* a concrete response to the Skolem–Gödel Antinomy of §4, in the specific case of the SET concept. It explains how creatures like us can acquire and manifest a SET concept which is so *precise*, that any purely set-theoretic claim is determinate.[35]

7. Internalism about model theory

I will say more about set theory in §9. But I now turn from set theory to model theory.

I dismissed *modelism* in §2. Now, it would certainly be a disaster, if this forced me to dismiss *model theory* itself.[36] Fortunately, it does not; my complaint against modelism is a complaint against a philosophical *misuse* of model theory. But this merits some explanation.

Modelists insist on using model theory to explicate the acquisition and manifestation of mathematical concepts. That is a mistake. Having recognised this, we must follow Putnam in "foreswear[ing] reference to models in [our] account of understanding" mathematical theories and concepts. But the modelist's mistake is no part of model theory *itself*, as a branch of pure mathematics. So, as Putnam notes, we do *not* "have to foreswear forever the notion of a model."[37] In fact, an internalist can introduce the MODEL concept quite easily.

In common with almost every branch of mathematics, model theory is largely carried out "informally": the proofs are discursive and they omit tedious steps. But we can easily make sense of the idea that, "officially", model theory is implemented within set theory. After all, model-theorists freely use set-theoretic vocabulary and set-theoretic axioms to describe and construct models, and all of the definitions from ordinary model theory could (in principle) be rewritten in entirely set-theoretic terms.

---

[34] See Button & Walsh (2018: §11.5).
[35] The approach, and invocation of a technical result, is similar to McGee (1997). For more on the similarities and differences, see Button & Walsh (2018: §11.A).
[36] Cf. Lewis (1991: 58) "To reject mathematics for philosophical reasons would be absurd."
[37] Putnam (1980: 179). See also Dummett (1963: 191, 193).

So, in what follows, $\mathrm{MT}_{\mathrm{int}}$ (for *M*odel *T*heory, *int*ernalised) will be a suitable set theory which I will use for model-theoretic purposes. There is no need to go into great detail about $\mathrm{MT}_{\mathrm{int}}$; I only need to explain how it relates to $\mathrm{CSP}_{\mathrm{int}}$. There are three crucial points:[38]

(1) $\mathrm{MT}_{\mathrm{int}}$ deals with a pure set property, P, and a membership relation, $\in$.
It might have other predicates too, but it has at least those.

(2) $\mathrm{MT}_{\mathrm{int}}$ proves $\mathrm{CSP}_{\mathrm{int}}$.
This means that $\mathrm{MT}_{\mathrm{int}}$ is internally categorical with respect to pure sets.

(3) $\mathrm{MT}_{\mathrm{int}}$ proves that there are infinitely many pure sets.
This gives $\mathrm{MT}_{\mathrm{int}}$ the resources to carry out basic reasoning concerning arithmetic and (arithmetized) syntax.

With these assumptions in place, $\mathrm{MT}_{\mathrm{int}}$ has all the basic vocabulary and conceptual resources for developing model theory as a branch of pure mathematics. Working model-theorists will almost certainly want to add more axioms to the underlying set theory, such as Replacement, but I will leave that to them (though I revisit the point in §8).

Internalists about model theory affirm $\mathrm{MT}_{\mathrm{int}}$ and insist that model theory is "officially" carried out deductively within $\mathrm{MT}_{\mathrm{int}}$. The payoff is as follows: via the internal categoricity and intolerance of $\mathrm{CSP}_{\mathrm{int}}$, we have an explanation of how we can acquire and manifest a MODEL concept which is so *precise*, that any purely model-theoretic claims are determinate.

(At some point, of course, we might well want to consider *impure* models, such as a model whose domain includes the cows in the field. However, this will not really affect much. The specifically model-theoretic features of an impure model can be determined by considering isomorphic models with pure domains.)

8. Internal realism, revisited

I have outlined internalist approaches to arithmetic, set theory and model theory. I now want to consider the interactions between internalism about these three branches of mathematics, with an aim to illuminating Putnam's mathematical internal realism.

Suppose that Sebastian has mastered arithmetic, in the form of $\mathrm{PA}_{\mathrm{int}}$, but that he knows no model theory. Still, we – who know some model theory – can pose a question:

Are any particular models of arithmetic "intended", from Sebastian's perspective?

The short answer is:

[38] If we want to develop an account of truth for $\mathrm{MT}_{\mathrm{int}}$ itself, then we should also insist that $\mathrm{MT}_{\mathrm{int}}$ be a single formula, so that we can continue to use it in the course of internal categoricity results, in the form of conditionals like $\forall P \forall \in (\mathrm{MT}(P, \in) \rightarrow \varphi)$. For details, see Button & Walsh (2018: §§12.4, 12.A).

Yes; Sebastian's use of PA_{int} *makes certain models "intended". This is because our model theory essentially proves that* PA_{int} *picks out a unique isomorphism type. So, it makes sense to say that the models of that type are "intended" for Sebastian.*

But I should spell this out carefully.

I will work within MT_{int}. For any number property, N, and any successor function, s, I will write $||N, s||$ for the model (as the notion is defined in MT_{int}) whose domain is the set whose members are exactly the instances of N, and whose interpretation of the successor-symbol is the set whose members are similarly determined by s.[39] Now suppose we augment MT_{int} with the principle "any countable property determines a set";[40] call the resulting theory MT_{int}^+.Then we can easily obtain:[41]

$$MT_{int}^+ \vdash \forall N_1 \forall s_1 \forall N_2 \forall s_2\, ([PA(N_1, s_1) \wedge PA(N_2, s_2)] \rightarrow \\ ||N_1, s_1|| \text{ is isomorphic to } ||N_2, s_2||)$$

Roughly, the point is that all PA_{int}-ish number concepts determine isomorphic models. So, within our model theory, we get to say: "the [deductive] use [of PA_{int}] already fixes the 'interpretation'" of PA_{int}.[42]

Now, the internalist about model theory *affirms* that model theory. So she affirms, unreservedly, that PA_{int} pins down a model up to isomorphism. And she can therefore *agree* with the modelist, of §2, that the NATURAL NUMBER concept is *precise up to isomorphism*.

In a sense, then, one *might* say that internalism employs "a similar picture" to modelism, only "within a theory".[43] *But this does not vindicate modelism itself.* For, to *show* that the NUMBER concept is precise up to isomorphism, the internalist works *within* model theory. And she claims to understand model theory *deductively*, rather than *semantically*.

This observation enables me to make sense of Putnam's cryptic but beautiful closing remarks in "Models and Reality". Since modelists always insist working *semantically*, they embark on a futile regress, and end up treating models as "lost noumenal waifs looking for someone to name them". (That was the point of §2.) However, by working *deductively*, internalists are able to treat models as "constructions within our theory itself, [which] have names from birth."[44] But saying this does *not* require any constructivist metaphysics. Rather,

---

[39] So, with standard abbreviations, $||N, s|| = (\{x : N(x)\}, \{\langle x, y\rangle : s(x) = y\}$

[40] This is easily formulable in second-order logic, and is entailed by e.g. second-order ZFC. This principle is needed, just because MT_{int} itself says *so little* about how many sets there are. So, MT_{int} itself proves a slightly weaker claim, which adds the conditional "provided all the required sets exist". However, I do not think this technical nicety materially much affects the philosophical point.

[41] This follows almost immediately from the internal categoricity of PA_{int}.

[42] Putnam (1980: 482).

[43] Putnam (1977: 484), commenting on how to regard the relationship between internal realism (in general) and metaphysical realism (in general).

[44] Putnam (1980: 482). Cf. also Dummett (1963: 191).

it simply rolls together some simple observations, namely that:[45] the definition of a model is offered *within* our deductively understood model theory; all talk of "construction" of models is just an heuristic shorthand for deductive work carried out *within* that model theory; and we work *within* our model theory when we prove that all models of PA_{int} are isomorphic, and when we say that Sebastian's use of PA_{int} picks out a particular isomorphism type.

The preceding few paragraphs can be summarised as follows: model theory itself does not *demand* a model-theoretic treatment. But, for exactly this reason, I expect the modelist to raise a complaint:

> **Modelist.** By working *semantically*, I can point out that MT_{int} has many models if it has any. And if you insist on only ever working *deductively*, then you cannot do anything to guard against the worry that we are "trapped in" a non-standard model of MT_{in} itself. But, you *must* rule out this worry, if you want to say that Sebastian (who knows no model theory) pins down the standard model by using PA_{int}. For, suppose we are all trapped in a non-standard model of MT_{int}. Then the claim "all PA_{int}-ish number properties determine the same model (if they determine one at all)" would still hold true for us. But what we happened to call the "intended model of arithmetic" might look grotesque, as viewed from the outside.[46]

Now, there is certainly *something* to this complaint. In discussing Sebastian's situation, I implicitly assumed that "the metalanguage", i.e. the model theory MT_{int}, "is completely understood" by us.[47] After all, if we could reasonably sharpen our MODEL concept in various different ways, then we might indeed find ourselves dealing with multiple different "models of arithmetic". And that would be a bad thing. Fortunately, MT_{int}'s intolerance precludes *precisely* this sort of situation: given rival sharpenings, only one of them can be *right*.

Furthermore, there is an easy reply to the modelist's worry. Suppose, for reductio, that we are "trapped" in some non-standard model, **M**, of MT_{int}. Working in MT_{int}, I can trivially prove: *every model's domain omits some element*. So now – if I can understand the modelist's worry that I am "trapped" in **M** *at all* – then I know, specifically, that **M**'s domain omits some elements. And if I can grasp *that* point, then I know that I am *not* "trapped" in **M** after all, since I just managed to quantify over the supposedly omitted elements.[48]

So: we are *not* "trapped in a non-standard model". But we should *not* infer from this that we "inhabit the *standard* model" of MT_{int}. Indeed, the same line of thought which shows that I am not "trapped" in **M** generalises to show that I do not "inhabit" *any* particular model of MT_{int}. Or, to drop the homely metaphors: *no* model of MT_{int} is the "intended" model.[49]

---

[45] Cf. Button (2013, 217).
[46] Cf. Meadows (2013, 539–40).
[47] Putnam (1980: 482).
[48] At this point, the modelist might start to say things like: *maybe I don't even understand (at all) the worry that I am "trapped" in **M**!* At that point, I feel we have earned the right to walk away from the modelist. Of course, there is more to say about such ineffable scepticism; but for more, see Button & Walsh (2018: ch.9, §11.6).
[49] Cf. Button & Walsh (2018: §11.6) on "indefinite extensibility" in this context.

If that initially sounds shocking, it really should not. Once we have abandoned modelism, there is no reason to think that our theory *needs* an "intended" model.

In this section, I have considered what internalists (in general) should say about models of PA_{int} and MT_{int}. The overarching moral is encapsulated in a single quote from Putnam: for any theory, "[e]ither the use already fixes the 'interpretation', or nothing can."[50] But I read this, not as a rhetorical flourish, but as a genuine disjunction.

In the case of PA_{int}, the use *already fixes* the interpretation. That is what we saw when we considered Sebastian.

In the case of MT_{int}, by contrast, we have seen that *nothing can*, for there *is* no intended interpretation. For all that, though, our model theory is not "uninterpreted syntax". We know how to use it – *deductively* – and our use manifests perfectly *precise* concepts. What more could we want or need?[51]

9. Coda

In this coda, I want to offer a few more remarks concerning how internalism might be further developed. This material is relegated to a mere coda, though, because it is *even more* speculative than the material in the previous sections. I am less committed to it, and I also suspect that aspects of it are genuinely *optional extras*.

Intersubjectivity, objectivity, and objects

One moral of §5 can be put as follows: *intolerance yields intersubjectivity*. More specifically: when a theory is intolerant, the parties who use that theory are not just deploying *private* concepts, but can (and must) be drawn into real (dis)agreement with each other. Still, whilst this yields a story about mathematical *intersubjectivity*, it delivers nothing (yet) about mathematical *objectivity*, or even about mathematical *objects*. I want to address that now. For simplicity, I will just focus on arithmetic, but I would say the same about sets.

The issue of *objects* can be settled almost immediately. As an internalist, I am committed to PA_{int}. I affirm it without reservation. And, in affirming it, I affirm that there *are* numbers. That is an end to it.

Moreover, this existential claim is crucial to the story of §5. There, I implicitly assumed that there are PA_{int}-ish number properties, i.e. that $\exists N \exists s\, \mathrm{PA}(N, s)$. After all, if that existential claim were *false*, then we would vacuously have that both $\forall N \forall s(\mathrm{PA}(N, s) \rightarrow \varphi)$ and $\forall N \forall s(\mathrm{PA}(N, s) \rightarrow \neg\varphi)$. Then, catastrophically, I would be forced to say, (for each relevant φ): it is determinate that φ *and also* determinate that $\neg\varphi$. So, my account of determinacy (and hence intersubjectivity) implicitly depends upon the existence of numbers.

---

[50] Putnam (1980: 482).

[51] Cf. Putnam's (1977: 489) "*Internal realism is all the realism we want or need.*"

To repeat, then: *I am committed to the existence of numbers*. But I have said very little about their *nature*. I have said that my numbers behave PA_{int}-ishly, but I have been silent about: whether the numbers are mind-independent or theory-independent; whether the number 2 is a Gallic emperor, or a set (and, if so, which); and, returning to the trivial illustration in §5, even whether the numbers are (capable of being) hungry.

I believe that I *could* say whatever I like about such matters. As such, I would really prefer to say *nothing at all*. Fortunately, there seems to be a principled way for an internalist to insist that all such matters are indeterminate.[52]

In §5, I glossed '$\forall N \forall s(\text{PA}(N, s) \rightarrow \varphi) \vee \forall N \forall s(\text{PA}(N, s) \rightarrow \neg\varphi)$' as 'it is determinate whether φ'. At the time, I restricted this gloss to sentences of a particular form (second-order formulas with only N and s free, and whose quantifiers are all restricted to N). But if I extend this gloss to cover sentences in *richer* languages, then I will get to say that it is indeterminate whether 2 is equal to Julius Caesar, or is hungry, or is (in)tangible. For if there are any PA_{int}-ish number properties, then there will be a number property which takes 2 to be a hungry, tangible, Caesar, and another which takes it to be a satiated, abstract, singleton set. More generally, on this approach, *all* questions about the "metaphysical nature" of numbers will have indeterminate answers. They can simply be ignored.

A "hardcore realist" might complain that this approach simply trivialises some *very important* questions in the metaphysics of mathematics.[53] So be it. My point is just that *internalists* get to say that all the facts about the numbers can be expressed in the language of arithmetic. And that strikes me, at least, as a nice additional point in favour of internalism.[54]

Conceptual relativity

I just discussed the significance of intolerance results; but I now want to come at their significance from a rather different angle.

In §5, I stated the significance of the intolerance result as follows: *If* Solange and Tristan share a logical language, *then* they just have to say "one of us as wrong", when one affirms Goldbach's Conjecture and the other affirms its negation. However, in what followed, I basically acted as if the antecedent is guaranteed to hold, without any further comment. But I should come clean: I have no guarantee that Solange and Tristan share a logical language. Moreover, if Solange and Tristan do not share a logical language, then in principle Solange might affirm φ, and Tristan might affirm $\neg\varphi$, and both of them could be right in their own languages.[55]

---

[52] This line is developed in Button & Walsh (2018: §10.7).

[53] The "hardcore realist" is a character from Putnam (1977: 490); thanks to Wesley Wrigley for suggesting I address this point.

[54] And it at least once sounded good to Putnam; see his comments on "whether the number 2 is identical with a set, and if so, which set is identical with" (1994: 248–51).

[55] Thanks to Cian Dorr, Hartry Field, and Luca Incurvati for discussion on this issue.

That said, it is hard to see how such a situation could arise. An immediate barrier to comprehension arises from my internalism; after all, the logical language is to be understood *deductively* rather than *semantically*, and we can assume that Solange and Tristan accept exactly the same rules of proof, so it is hard to see what it could mean to say that they *do not* share a logical language.

Still – and this is *very* speculative – I might *just* be able to illustrate the possibility, by drawing an analogy with Putnam's discussions of mereology.[56]

Imagine two characters, Stan and Rudy. Stan is a mereological universalist, and thinks that *any things compose a fusion*. Rudy is a nihilist, and thinks that *there are no fusions*. Stan and Rudy might argue vociferously about which of them *gets the world right*. But at least one reasonable response to their dispute is to see them not as *disagreeing*, but as operating with *different conceptual schemes* (or frameworks, or languages, or whatever). This response is reinforced by the idea – which Putnam affirmed – that we can translate back and forth between Stan and Rudy's ways of talking. Roughly: Stan is to interpret all of Rudy's quantifiers as restricted to what Stan calls "simples"; Rudy is to interpret Stan's talk of "fusions, composed of simples" as talk of "plurals, among which there are simples".

On the specific issue of mereology, the devil will be in the details. But the details about mereology can be set aside for now. At a high level of description, the thought is just this: Rudy and Stan can offer deviant interpretations of each other's "logical concepts", and thereby dissolve their apparent disagreement.

Returning to the case of arithmetic: I can see no reason *in principle* why a similar thing might not happen with Solange and Tristan. If they apparently disagree, we might (for all I know) be able to give them a suitable translation manual which smooths over the difference. And, *in principle*, that might be the right thing to do. But I emphasise: *in principle*. Rudy and Stan are equally successful in navigating their way around the world. Confronted with the same situation, they systematically give different – but wholly predictable – answers to the question "how many things are there?" So it is *deeply* reasonable to think that they are simply speaking different languages; just using different words in the same situations. It is vastly harder to see what would prompt a similar thought in the arithmetical case. Truth told, I cannot think of *anything*. But maybe this is just lack of imagination on my part.

Having said all this, I should close with a simple, but important, point: admitting the (in principle) possibility of reinterpreting logical vocabulary is entirely compatible with my earlier discussion of objects and objectivity. Tolerance concerning reinterpretation "is not a facile relativism that says 'Anything goes'."[57] I am allowing only that we might have some freedom to choose between two languages, such that φ is the right thing to say in one language and $\neg\varphi$ is the right thing to say in the other. But if Solange has fixed a language and affirms φ, and if Tristan now affirms $\neg\varphi$, then Solange must *either* regard Tristan as saying

[56] See in particular Putnam (1987), and Button (2013, chs.18–19).

[57] Putnam (1981: 54).

something false, *or* regard Tristan as speaking a different language. This disjunction yields no sacrifice of objectivity, for it is entirely commonplace. If Tristan says "I have a pet pink elephant", I have the same two options: I must *either* regard him as saying something false, or regard his words as expressing something other than my own; but this does not make it "up to me" whether any pink elephants *exist*.

In summary, then, here is how I should cautiously state the significance of an intolerance theorem. In principle, I am tolerant when it comes to choosing languages. But, *within* a language and in the presence of an intolerance theorem, divergence cannot be tolerated. Moreover – given what I just said about pink elephants – this level of intolerance secures all the objectivity – and all the realism – that anyone could either want or need.

The continuum hypothesis

Internalism about arithmetic delivers the verdict that every arithmetical claim is *determinate*. That is one of its main virtues. However, internalism about set theory *also* delivers the verdict that every pure-set-theoretical claim is determinate. And this is less clearly a *virtue*. No doubt many people will reply that it *surely* cannot be so easy, to arrive at the conclusion that the continuum hypothesis is determinate (for example).

I fully feel the force of this concern. But I will close by saying a few things, to try to diminish its force a little.

First: I am only claiming that the continuum hypothesis *is* determinate. I am not suggesting that we will ever be able to *know* whether it holds. Indeed, the existence of unknowable mathematical truths is perfectly compatible with internalism.

Second: the Intolerance Theorem for CSP ineliminably requires a deductive system for *impredicative* second-order logic.[58] So: maybe those who think that the continuum hypothesis is indeterminate should reject *impredicative* reasoning. If so, though, that would be a deeply interesting connection, but I cannot purse it here any further.

Third: despite everything I have said, there may yet be a way to accept impredicative reasoning whilst making some sense of the "indeterminacy of CH". The possibility arises given the speculative discussion of the previous subsection. There, I nuanced my claims about the intolerance of arithmetic, and I should nuance my claims about the intolerance of set theory in the same way. So: once you have *fixed* a logical language, the claim "there is no cardinal between the cardinality of the naturals and the cardinality of the reals" becomes *determinate* (if not *decided* by the theory); but, *in principle*, different logical languages may settle it differently. And perhaps the possibility of tolerance in *choosing a logical language* is all that is needed, for those who want to explore the "(in)determinacy of CH".[59]

---

[58] For a demonstration of the need for impredicativity, see Button & Walsh (2018: §11.C)

[59] I wrote this paper during a period of research leave which was funded by a Philip Leverhulme Prize (awarded by the Levehulme Trust, PLP–2014–140). For comments and discussions on drafts of this paper (and earlier versions of these thoughts), I would like to thank Neil Barton, Sharon Berry, Cain Dorr, Hartry Field, Luca Incurvati, Mary Leng, Guy Longworth, Vann McGee, Charles Parsons, Michael Potter, Robert Trueman, Sean

References

Benacerraf, Paul. 1985. 'Skolem and the skeptic'. *Proceedings of the Aristotelian Society* 59, 85–115.

Button, Tim & Sean Walsh. 2018. *Philosophy and Model Theory*. Oxford: Oxford University Press.

Button, Tim. 2013. *The Limits of Realism*. Oxford: Oxford University Press.

Dummett, Michael. 1963. 'The philosophical significance of Gödel's Theorem'. Page references to Dummett, *Truth and Other Enigmas*. 1978, London: Duckworth.

Florio, Salvatore & Øystein Linnebo. 2016. 'On the innocence and determinacy of plural quantification'. *Noûs* 50.3: 565–83.

Gödel, Kurt. 1951. Some basic theorems on the foundations of mathematics and their implications. Page references to *Kurt Gödel: Collected Works, III*. 1995. Edited by S. Feferman et al. Oxford: Oxford University Press. 304–23.

Lewis, David. 1991. *Parts of Classes*. Oxford: Blackwell.

McGee, Vann. 1997. 'How we learn mathematical language'. *Philosophical Review* 106.1, 35–68.

Parsons, Charles. 1990. 'The uniqueness of the natural numbers'. *Iyyun* 39.1: 13–44.

Parsons, Charles. 2008. *Mathematical Thought and Its Objects*. Cambridge: Harvard University Press.

Potter, Michael. 2004. *Set Theory and its Philosophy*. Oxford: Oxford University Press.

Putnam, Hilary. 1977. 'Realism and reason'. *Proceeds of the American Philosophical Society* 50.6: 483–98.

Putnam, Hilary. 1980. 'Models and reality'. *The Journal of Philosophical Logic* 45.3: 464–82.

Putnam, Hilary. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

Putnam, Hilary. 1987. 'Truth and convention'. *Dialectica* 40.1–2: 69–77.

Putnam, Hilary. (2000). 'Das modelltheoretische Argument und die Suche nach dem Realismus des Common sense'. In *Realismus*. Edited by M. Willaschek. Paderbon: Ferdinand Schoningh Verlag: 125–42.

Rips, Lance J, Amber Bloomfield & Jennifer Asmuth. 2008. 'From numerical concepts to the concept of number'. Behavioral and Brain Science 31.6: 623–87.

Väänänen, Jouko, and Tong Wang (2015). 'Internal categoricity in arithmetic and set theory'. *Notre Dame Journal of Formal Logic* 56.1: 121–34.

Wittgenstein, Ludwig. 1956. *Remarks on the Foundations of Mathematics*. Translated by G E M Anscombe. Blackwell: Oxford.

Walsh, Crispin Wright, and Wesley Wrigley. But I would like to thank Sean Walsh most of all: we wrote *Philosophy & Model Theory* together, and this paper owes *so* much to our collaboration there.