Tim Button
Mathematical Internal Realism

Abstract: In "Models and Reality" (1980), Putnam sketched a version of his internal realism as it might arise in the philosophy of mathematics. Here, I will develop that sketch. By combining Putnam's model-theoretic arguments with Dummett's reflections on Gödelian incompleteness, we arrive at (what I call) the Skolem-Gödel Antinomy. In brief: 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. Whilst this might initially seem strange, I show how internal categoricity results for arithmetic and set theory allow us to face up to this Antinomy. This also allows us to understand why "Models are not lost noumenal waifs looking for someone to name them," but "constructions within our theory itself," with "names from birth."

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 paper 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 footnotes. But I should like to make it perfectly clear that, without the work of the people just mentioned, this paper could not even have begun.

I Acquisition and Manifestation

I want to start by considering our *natural number* concept. For clarity: I am not interested in specific number concepts, like *four* 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 baby has the general *number* concept.

Equally, we must be able to manifest our mathematical concepts. Whilst mathematicians 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 (fish, fingers, beads, or cows). Later, we master algorithms for adding and multiplying numbers in decimal notation. And so it goes. 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 and pedagogical steps we might take towards acquiring that concept, we qualify as having acquired it fully, only when we have grasped some full-blown arithmetical theory, such as Peano arithmetic.[2] Equally, we fully manifest our grasp of the concept, only by articulating and using some such theory.

In what follows, then, I will assume that serious mathematical concepts can be (and only can be) fully acquired and manifested by mastering and articulating some theory. Much more could be said in defence of this assumption. But I think the assumption is correct, and this paper is an attempt to work through its consequences. In §§II–IV, I will explain how this assumption threatens to constrain the precision of our mathematical concepts; then, in §§V–X, I will explain how we can overcome that threat by developing Putnam's internal realism.

---

1 Here I intend to connect with Dummett's long-held insistence on the importance of manifestation and acquisition (see e.g., Michael Dummett, "The Philosophical Significance of Gödel's Theorem" (1963), reprinted in *Truth and Other Enigmas* (London: Duckworth, 1978), 188–190).

2 For interesting discussion concerning the stage at which we (implicitly) grasp Peano arithmetic (or something like it), see Lance J. Rips, Amber Bloomfield and Jennifer Asmuth, "From Numerical Concepts to the Concept of Number," *Behavioral and Brain Science* 31, no. 6 (2008): 623–687 and the subsequent "Open Peer Commentary."

II Modelism

Consider this question: *How precise is our natural number concept?*[3] A specific philosophical character, the *modelist*, answers this question with a slogan. She says:

> The *natural number* concept is precise up to isomorphism.

But, of course, the modelist will need to flesh out this slogan. To this end, she makes the following speech:

> 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*.[4] And that is what I mean, when I say that the *number* concept is precise up to isomorphism. I mean that we can (and should) use an isomorphism type as a surrogate for the *number* concept.

> Note that many mathematical concepts are not so 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.

---

3 Dummett and Charles Parsons ask roughly this question (see Dummett's "The Philosophical Significance of Gödel's Theorem," and Parsons' "The Uniqueness of the Natural Numbers," *Iyyun* 39, no. 1 (1990): 13–44). Hilary Putnam raises very similar issues, but via questions which focus more on *objects* than on *concepts* ("Models and Reality," *The Journal of Philosophical Logic* 45, no. 3 (1980): 464–482). However, objectual and conceptual versions of the question are very similar (see Tim Button and Sean Walsh, *Philosophy and Model Theory* (Oxford: Oxford University Press, 2018), Chs. 6–8). So, for simplicity, I will focus solely on the conceptual version.

4 Our modelist might do better to focus on definitional equivalence instead of isomorphism (see Button and Walsh, *Philosophy and Model Theory*, §§5.1–5.2); but this would not change the dialectic, so I will ignore this complication.

That is modelism, in a nutshell. Modelism is obviously structuralist, but it is just one version of structuralism. And its special reliance on model theory gives rise to its name, *modelism*.[5]

Modelism is appealing. Unfortunately, as Putnam taught us, it is dead wrong. It succumbs to the *model-theoretic argument*.[6]

In §I, I insisted that mathematical concepts must be tied to theories, via manifestation and acquisition. So, if the modelist is right that the *number* concept is precise up to isomorphism, 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 a unique isomorphism type. In which case, 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 §I.

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*. As a practice, arithmetic is not just a list of axioms, but rather a "MOTLEY of techniques and proofs," to use Wittgenstein's

---

5 Button and Walsh coined the term "modelism"; see *Philosophy and Model Theory*, Ch. 6 for more.

6 The remainder of this section presents the central problem I extract from Putnam's invocation of the Löwenheim-Skolem Theorem in "Models and Reality." (Admittedly, Putnam raised the issue in a more "objectual" than "conceptual" key; but see footnote 3, above.) Dummett raised a similar problem, focussing on Gödelian incompleteness (see "The Philosophical Significance of Gödel's Theorem," 192). For more, see Button and Walsh, *Philosophy and Model Theory*, Ch. 7.

imagery.[7] A modelist might want to suggest that this informal motley plays some role in picking out an isomorphism type.[8]

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, we cannot just deploy results from model theory to raise problems for her. And this might seem like a strike in favour of an "informalist" modelism.

However, this point 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 a unique isomorphism type. Moreover, leaving this issue unexplained is not a viable option. After all, to treat the matter as inexplicable would be to say that it is just a brute feature of the world – a "surd metaphysical fact"[9] – that our informal mathematical practice pins down one particular isomorphism type. And this would be tantamount to the patently ridiculous claim:

> Everyone who wears this *particular* motley just happens to pick out this very *specific* thing; which is really rather fortunate, since (a priori) any of us might have picked out different things, or indeed have failed to pick out anything at all!

On pain of embarrassment, then, I take it that modelists *are* restricted to using formal theories, and will seek to explain how such theories can pin down isomorphism types.[10]

As I presented the model-theoretic argument, though, I did not just assume that the modelist's favourite theory must be formal; I also assumed that the theory must be first-order (and countable). To explain why this is a significant assumption, allow me to mention some simple technicalities. 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. (This allows us to gloss "$\forall X$"

---

7 Ludwig Wittgenstein, *Remarks on the Foundations of Mathematics*, translated by G.E.M. Anscombe (Oxford: Blackwell, 1956), §46.

8 This seems to be Benacerraf's response to Putnam's "Models and Reality." See Paul Benacerraf and Crispin Wright, "Skolem and the Skeptic," *Proceedings of the Aristotelian Society 59* (1985): 108–111.

9 To use Putnam's phrase from Hilary Putnam, *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981), 48. (The context of the quote is the permutation argument against metaphysical realism in general; but the same thought applies here.)

10 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 consider informal practice. But this does not affect the general point that, in terms of §1, the concept is manifested with its full precision by (and only by) use of a formal theory.

roughly as "for any subset of the first-order domain.") Neither the Löwenheim-Skolem nor the Compactness theorems hold, given this semantics. On the contrary, we have this:[11]

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

So, the modelist might reply to the model-theoretic argument by invoking Dedekind's result, and saying:

> 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 it is fatally flawed.[12] The flaw does not concern the use of *second-order* Peano arithmetic; there is nothing intrinsically wrong with allowing quantification into predicate-position. 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 her favourite 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 her to reply as follows:[13]

> The key concept, i.e. *powerset*, is just the concept of *all combinatorially possible subcollections of a collection*.

---

11 For a modern proof, and references to plenty of other proofs, see e.g. Button and Walsh, *Philosophy and Model Theory*, §7.4.

12 What follows is, in effect, one version of Putnam's famous just-more-theory manoeuvre. See Hilary Putnam, "Realism and Reason," *Proceedings of the American Philosophical Society* 50, no. 6 (1977): 486–487 and "Models and Reality," 477, 481, as well as references and discussion in Tim Button, *The Limits of Realism* (Oxford: Oxford University Press, 2013), Chs. 4–7 and Button and Walsh, *Philosophy and Model Theory*, §§2.3, 7.7–8.

13 Thanks to Mary Leng for suggesting this way of putting it.

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 §I apply.[14]

In §I, 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 before: 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.

(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 being 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.[15] Equally clearly, it is vicious. It simply cannot be a constraint, on acquiring or manifesting the concepts involved in one

---

14 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? Salvatore Florio and Øystein Linnebo, "On the Innocence and Determinacy of Plural Quantification," *Noûs* 50, no. 3 (2016): 565–583 develop this criticism elegantly.

15 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.

theory, that we must first acquire or manifest the concepts involved in the theory at the next level; if it were, then we would never be able to acquire or 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 attempted to rebut the model-theoretic argument by invoking any of several alternative logics. But there is a hard limit on this strategy. As noted above, the Compactness Theorem is sufficient to yield a model-theoretic argument. But Compactness holds for any logic with a finitary (sound and complete) proof system.[16] 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 proof-theoretically, but must instead be articulated semantically. And that suffices to set the modelist off on her vicious regress.

III A Dummettian Approach

Modelism has failed. We need an alternative. The obvious thought is simply to try approaching matters proof-theoretically, rather than model-theoretically. Indeed, this was Dummett's approach. His central idea can be stated as follows:

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

This idea promises to handle the requirements of acquisition and manifestation better than modelism did. After all, when it comes to teaching and learning mathematics, rules of proof are rather more tractable than isomorphism types.

Unfortunately, there is an immediate barrier to this proposal. Let P be any algorithmically-checkable proof-system, by which I mean that there is an algorithm which decides whether any putative P-proof is a genuine P-proof. Now, suppose for reductio that P-provability exhausts the arithmetical facts, i.e., that, for every arithmetical sentence φ:

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

Since our proof-system is algorithmically-checkable, some computable function captures the idea that n is (the code of) a P-proof of (the code of) φ. This means that there will be an arithmetical predicate, Tr, such that, for any arithmetical sentence φ:

---

16 See Button and Walsh, *Philosophy and Model Theory*, §7.9.

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

Combining the biconditionals, for any arithmetical sentence φ:

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

But this contradicts Tarski's Indefinability Theorem.[17] So *P*-provability does not exhaust the arithmetical facts after all. Generalising on *P*, we obtain:

(b)   No algorithmically-checkable proof-system exhausts the arithmetical facts.

Dummett is aware of this sort of reasoning,[18] but he does not take it to undermine (a). Instead, he *ponenses* where others might *tollens*. Since Dummett 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."[19] That is, combining (a) with (b), he concludes that that the *number* concept itself "cannot be fully expressed by means of any formal system."[20]

　　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 manifest and acquire concepts which *can* be fully expressed by means of some formal system.[21] 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 itself, but can only acquire some imprecise approximation to it. In short, Dummett is committed to a startling disjunction:

(c)   Either we are not machines, or we do not possess the *number* concept.[22]

---

17 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 a notion of arithmetical truth that is (somehow) "prior" to a notion of proof.

18 Though Dummett, "The Philosophical Significance of Gödel's Theorem," focusses on Gödelian reasoning, rather than on Tarskian undefinability.

19 Dummett, "The Philosophical Significance of Gödel's Theorem," 200.

20 Dummett, "The Philosophical Significance of Gödel's Theorem," 186.

21 Setting aside machines with access to oracles.

22 This is obviously similar to Gödel's Disjunction (Kurt Gödel, "Some Basic Theorems on the Foundations of Mathematics and Their Implications" (1951), reprinted in Kurt Gödel, *Collected Works*, Vol. 3, edited by Solomon Feferman et al. (Oxford: Oxford University Press, 1995), 310). However, the right disjunct here ("we do not possess the *number* concept") should be contrasted with Gödel's ("there exist absolutely unsolvable diophantine problems").

I cannot take seriously the possibility that we do not possess the *number* concept. Equally, though, I cannot allow that our philosophy of mathematics might require that we are not machines. I therefore have no option but to part ways with Dummett.

IV The Skolem-Gödel Antinomy

The previous three sections can be summarised as follows. We (fully) acquire and manifest our mathematical concepts via formal theories. Modelism treats such theories model-theoretically. In so doing, it succumbs to Putnam's model-theoretic argument. The obvious alternative is to treat formal theories proof-theoretically. But, to allow for the possibility that we are machines, the relevant proof-system must be algorithmically-checkable; and the *number* concept is sufficiently precise that no algorithmically-checkable proof-system exhausts the arithmetical facts. All told, then, we find ourselves in the following predicament:

The Skolem-Gödel Antinomy. Our mathematical concepts are perfectly precise. However, these perfectly precise mathematical concepts are (fully) acquired and manifested via a formal theory, which is understood in terms of an algorithmically-checkable proof-system, and hence is incomplete.

Confronted with this antinomy, one might well worry that *something must have gone wrong*: *surely* any concept which is (fully) articulated in an *incomplete* theory must be *imprecise*? I certainly feel the tension; indeed, that is why I call this predicament an "antinomy."[23] Still, I do not think that anything has gone wrong. This really is our predicament, and we must 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, as I will show, internalism amounts to a detailed development of the mathematical internal realism which Putnam sketched at the end of his "Models and Reality."[24]

---

23 Cf. Putnam's ("Models and Reality," 464) use of "antinomy."

24 The material in the second half of this paper develops joint work with Sean Walsh (Button and Walsh, *Philosophy and Model Theory*, Chs. 10–12). In that work, Sean and I did not endorse internalism; we simply wanted to articulate the best possible version of internalism. In this paper, I want to stick my neck out slightly further. Here is how.

I am confident that the Skolem-Gödel Antinomy accurately describes our predicament. Moreover, internalism strikes me as the most promising line of response to that Antinomy. Indeed, at the moment, I see no other way to face up to the Antinomy. Still, there is much more work to be

V 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 to assume that everything is a number. So I need a primitive predicate, "$N(x)$" which is 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 also introduce two obvious abbreviations:

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

Using these symbols and abbreviations, I can lay down four axioms:

(1) $(\forall n: N) N(s(n))$
 i.e. the successor of any number is a number
(2) $(\exists z: N)(\forall n: N) s(n) \neq z$
 i.e. there is a zero-element
(3) $(\forall m: N)(\forall n: N)(s(m) = s(n) \rightarrow m = n)$
 i.e. successor is injective on the numbers
(4) $\forall F([(\forall z: N)((\forall n: N) s(n) \neq z \rightarrow F(z)) \wedge (\forall n: N)(F(n) \rightarrow F(s(n)))] \rightarrow (\forall n: N) F(n))$
 i.e. induction holds for the numbers: 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*ernalised, since PA_{int} is just second-order Peano arithmetic, with all the axioms relativised to "N". This is the theory which I will wield in the face of the Skolem-Gödel Antinomy.

To appreciate the virtues of PA_{int}, imagine that Solange and Tristan have both learned PA_{int}.[25] They are now happily babbling away to each other, exploring the

---

25 The idea here is inspired by Parsons's discussions of Kurt and Michael in "The Uniqueness of the Natural Numbers." For more on the similarities and differences between this approach and Parsons's, see Button and Walsh, *Philosophy and Model Theory*, §10.B.

theory's consequences. They shared a teacher, and so they use the same word-types as each other. Still, 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, and 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 of course 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 (her tummy rumbling) can rightly say "zero is hungry," whilst Tristan (satiated) 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 about whether the numbers are hungry. We philosophers should probably do the same.

It is, then, unreasonable to ask for a guarantee that Solange and Tristan are talking about the same objects. It is much more reasonable to ask for a guarantee that Solange's numbers and Tristan's numbers share the same arithmetical structure. And the following result provides just such a guarantee:[26]

Internal Categoricity of PA

$$\vdash \forall N_1 \forall s_1 \forall N_2 \forall s_2([PA(N_1, s_1) \wedge PA(N_2, s_2)] \rightarrow$$
$$\exists R\, [\forall v \forall y(R(v, y) \rightarrow [N_1(v) \wedge N_2(y)]) \wedge$$
$$(\forall v\colon N_1) \exists! y R(v, y) \wedge$$
$$(\forall y\colon N_2) \exists! v R(v, y) \wedge$$
$$\forall v \forall y(R(v, y) \leftrightarrow R(s_1(v), s_2(y)))])$$

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:

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

---

26 See Button and Walsh, *Philosophy and Model Theory*, §10.B, and Juoko Väänänen and Tong Wang, "Internal Categoricity in Arithmetic and Set Theory," *Notre Dame Journal of Formal Logic* 56, no. 1 (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-variable, leaving it to context to individuate what treatment is appropriate. For a rigorous treatment, see Button and Walsh, *Philosophy and Model Theory*, Chs. 10–12.

This *internal* categoricity result evidently resembles Dedekind's categoricity result, that all models of second-order Peano arithmetic are isomorphic (see §II). But it is worth spelling out the deep 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. It is proved within the ordinary deductive system for (impredicative) second-order logic (as indicated by the use of '⊢' in the statement of the Internal Categoricity Theorem). The proved sentence is in the same language as PA_{int} itself (indeed, it is a sentence of the "purely logical" fragment of PA_{int}). So it is an internal categoricity theorem, in precisely the following sense: it neither takes us beyond the object language, nor outside that language's proof-system.

Sticking with deduction has a benefit. In §II, our modelist attempted to invoke Dedekind's categoricity result. This forced her to insist that some particular semantic theory was privileged, and this set her off on a vicious regress. Since the Internal Categoricity Theorem invokes no semantic notions, no similar regress can arise.

However, sticking with deduction also has a cost. Inevitably, PA_{int} cannot prove its own Gödel-sentence. Since we are viewing PA_{int} deductively, we must therefore see it as *incomplete*.

Such incompleteness was, of course, promised to us by the Skolem-Gödel Antinomy. Nonetheless – and to address that Antinomy – I now want to explain why PA_{int} succeeds in introducing a perfectly precise *number* concept.

Let us revisit Solange and Tristan, respectively affirming $PA(N_1, s_1)$ and $PA(N_2, s_2)$. Suppose that Solange affirms (an appropriate formalisation of) "every even number is the sum of two primes," at which Tristan shakes his head and replies "some even number is not the sum of two primes." Now, we already noted that Solange and Tristan need not agree about what the numbers are. Still, we might hope that Solange and Tristan are genuinely disagreeing here, rather than merely talking past each other in their different languages. After all, Goldbach's Conjecture is purely arithmetical, and all of Solange's arithmetical structure is mirrored in Tristan's numbers, and vice versa, so, *surely* they are genuinely engaged with each other?

Indeed they are. This follows from a neat corollary of PA_{int}'s internal categoricity:[27]

---

27 For a full statement and proof, see Button and Walsh, *Philosophy and Model Theory*, §§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 and Walsh, *Philosophy and Model Theory*, §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:

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

So, when Solange affirms Goldbach's Conjecture whilst Tristan denies it (in their respective languages), 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 they must hold that one of them is *wrong*; for Goldbach's Conjecture must hold of all PA_{int}-ish *number* concepts (the left disjunct of the Intolerance Theorem) or fail of all of them (the right disjunct).

More generally, we can gloss the Intolerance Theorem as follows:

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

This explains why I call the result an *intolerance* theorem; it shows that PA_{int} does not tolerate different ways of pursuing arithmetic.

The Intolerance Theorem underpins my claim that PA_{int} articulates the *number* concept precisely. To spell out the last steps towards this conclusion, 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* PA_{int}-ish *number* concept, then we cannot render φ wrong, just by considering Tristan's *number* concept rather than Solange's, or whatever. So, by the above heuristic, it is determinate that φ. More generally, this suggests that we should gloss $\forall N \forall s(\mathrm{PA}(N, s) \to \varphi)$ 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:

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

i.e.:   *either it is determinate that φ or it is determinate that ¬φ*
i.e.:   *it is determinate whether φ*

In sum: thanks to its intolerance, PA_{int} articulates our *natural number* concept sufficiently precisely, that *every arithmetical claim is determinate*.

Allow me to summarise this section. The theory PA_{int} can be stated very briefly – it has just four conjuncts – so 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 PA_{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 §IV, in the specific case of the *number* concept. That is the response I want to offer. And here is a more general statement of internalism (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 an algorithmically-checkable proof-system. 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*.

VI Intersubjectivity, Objectivity, and Objects

One moral of §V can be put as follows: *intolerance yields intersubjectivity*. More specifically: when a theory is intolerant, people using that theory are not just deploying *private* concepts, but are drawn into genuine (dis)agreement with each other. So, internalism provides an account of mathematical *intersubjectivity*. It is worth, though, briefly connecting this with issues about mathematical *objectivity* and mathematical *objects*.

As an internalist, I am committed to PA_{int}. I affirm it without reservation. And, in affirming it, I affirm that there *are* numbers: the axioms carry existential commitment. (I should be frank, and admit that I am not sure exactly how best to answer the question: *How do you know that there are numbers?* Still, as an internalist, I am committed to their existence.)

Moreover, this existential commitment is indispensable to the story I told in §V. To see why, suppose that there were no PA_{int}-ish number properties, i.e. that $\neg\exists N\exists s PA(N, s)$. Then we would vacuously have that both $\forall N\forall s(PA(N, s) \rightarrow \varphi)$ and $\forall N\forall s(PA(N, s) \rightarrow \neg\varphi)$, for each relevant φ. It would follow that that it is both determinate that φ and determinate that $\neg\varphi$. This would be catastrophic. So, contraposing, the satisfactoriness of my account of determinacy (and hence intersubjectivity) implicitly requires that $\exists N\exists s PA(N, s)$.

To repeat, then: internalists are 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 many things: about whether the numbers are mind-independent or theory-independent; about whether any number is a Gallic emperor, or a set (and, if so, which); and, returning to the trivial illustration in §V, even about whether the numbers are hungry.

I believe that I could say whatever I like about such matters. For this reason, I would really prefer to say nothing at all. It is fortunate, then, that there is a principled way for an internalist to insist that *all such matters are indeterminate.*[28]

In §V, I glossed $\forall N \forall s(\text{PA}(N, s) \rightarrow \varphi) \lor \forall N \forall s(\text{PA}(N, s) \rightarrow \neg\varphi)$ 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 the number 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 an 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 brisk response simply trivialises some *very important* questions in the metaphysics of mathematics.[29] Let them complain. 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 as a nice "bonus point" in favour of internalism.[30]

VII Internalism about Set Theory

There is much more to say about internalism about arithmetic. I will say some of it in §X. First, I want to consider internalism about set theory. In brief, I want to lift the story of §§V–VI over from the *number* concept to the *set* concept.

As in §V, I will start by introducing an "internalised" theory of pure sets. Rather than using a Zermelo-Fraenkel-style theory, though, I prefer to use a set theory which captures the "minimal core" of the cumulative iterative notion of

---

28 This line is developed in Button and Walsh, *Philosophy and Model Theory*, §10.7.

29 The "hardcore realist" is a character from Putnam "Realism and Reason," 490. Thanks to Wesley Wrigley for suggesting I address this point.

30 Something similar also 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." ("Comments and Replies," in *Reading Putnam*, edited by Peter Clark and Bob Hale (Oxford: Blackwell, 1994), 248–251)

set. This *Level Theory* has its origins in work by Montague, Scott, Derrick, and Potter.[31]

I want to articulate a theory of pure sets. This is not to say that there is no set of the cows in the field; only that (for present purposes) I will ignore that set if it exists. 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, "$\in$." Using these symbols and the abbreviations of §V, I can then write down some axioms:

(1) $\forall x \forall y (x \in y \rightarrow (P(x) \land P(y)))$
    i.e. we restrict our attention to membership facts between pure sets
(2) $(\forall x: P)(\forall y: P)[\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y]$
    i.e. pure sets are extensional entities
(3) $\forall F(\forall x: P)(\exists y: P)\forall z(z \in y \leftrightarrow (F(z) \land z \in x))$
    i.e. pure sets obey separation
(4) $(\forall x: P)(\exists v: Level)x \subseteq v$
    i.e. every set is found at some level

As written, principle (4) uses an undefined predicate, "*Level*". However – and this is the neat trick about the approach – we can explicitly define "*Level*" in terms of set-membership.[32] As such, the only primitives we need are "*P*" and "$\in$." Let LT_{int} (for *L*evel-*T*heory, *int*ernalised) be the conjunction of these four axioms.

---

31 I present a simple version of the theory in Tim Button, "Level Theory, Part 1: Axiomatizing the Bare Idea of a Cumulative Hierarchy of Sets," in *Bulletin of Symbolic Logic* 27, no. 4 (2021): 436–460. For the origins of Level Theory, see: Richard Montague, "Set Theory and Higher-Order Logic," in *Formal Systems and Recursive Functions. Proceedings of the Eight Logic Colloquium, July 1963*, edited by John Crossley and Michael Dummett, 131–148 (Amsterdam: North-Holland, 1965); Richard Montague, Dana Scott, and Alfred Tarski, "An Axiomatic Approach to Set Theory" (BANC MSS 84/69c, carton 4, folder 29–30, Bancroft Library, University of California, Berkeley); Dana Scott, "The Notion of Rank in Set-Theory," in *Summaries of Talks Presented at the Summer Institute for Symbolic Logic, Cornell University, 1957*, 267–269 (Princeton: Institute for Defence Analysis, 1960); Dana Scott, "Axiomatizing Set Theory," in *Axiomatic Set Theory II. Proceedings of the Symposium in Pure Mathematics of the American Mathematical Society, July-August 1967,* 207–214 (Providence: American Mathematical Society, 1974) and Michael Potter, *Set Theory and Its Philosophy* (Oxford: Oxford University Press, 2004), especially Ch. 3. Thanks to Charles Parsons for making me aware of Montague's work. For a brief presentation of all that is required for the purposes of this paper, see Button and Walsh, *Philosophy and Model Theory*, §§8.B–C, 11.C–D.
32 For the definition, see Button, "Level Theory, Part 1," Definition 2.2. This simplifies a definition due to Potter, *Set Theory and Its Philosophy*, 24, 41, which is also used in Button and Walsh, *Philosophy and Model Theory*, §§8.5, 8.C.

Crucially, LT_{int} proves that the levels are well-founded by membership. This is why LT_{int} provides 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 for a moment, (the pure parts of) the full second-order models of LT_{int} are, up to isomorphism, exactly the arbitrary stages of the (pure) cumulative hierarchy of sets, as axiomatised by second-order ZF.[33] But I mention this fact, only to make LT_{int} feel a bit more familiar. I will treat LT_{int} deductively, just as I treated PA_{int} in §V.

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 LT_{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,[34] and skip straight to a theory which is internally (totally) categorical. I call this theory CLT_{int}, for *Categorical Level-Theory*. We obtain it by adding a fifth conjunct to LT_{int} (where "f" is a second-order function-variable):

(5) $\exists f (\forall x P(f(x)) \wedge \forall y(P(y) \rightarrow \exists! x\, f(x) = y))$
     i.e. there are exactly as many pure sets as there are objects simpliciter
     (i.e. objects which are either pure sets or not).

In the deductive system for impredicative second-order logic, we can then prove internal categoricity for CLT_{int}. Informally, this says that there is a membership-preserving bijection from Solange's pure sets to Tristan's. Formally:[35]

Internal Categoricity of CLT.
$\vdash \forall P_1 \forall \in_1 \forall P_2 \forall \in_2 ([CLT(P_1, \in_1) \wedge CLT(P_2, \in_2)] \rightarrow$
$\quad\quad \exists R\, [\forall v \forall y(R(v, y) \rightarrow [P_1(v) \wedge P_2(y)]) \wedge$
$\quad\quad\quad (\forall v{:}P_1)\exists! y R(v, y) \wedge$
$\quad\quad\quad (\forall y{:}P_2)\exists! v R(v, y) \wedge$
$\quad\quad\quad \forall v \forall x \forall y \forall z([R(v, y) \wedge R(x, z)] \rightarrow [v \in_1 x \leftrightarrow y \in_2 z])])$

---

33 See Button and Walsh, *Philosophy and Model Theory*, §8.C.

34 Interested readers should look to Button and Walsh, *Philosophy and Model Theory*, §§11.2, 11.C and Button, "Level Theory, Part 1", §6.

35 See Button and Walsh, *Philosophy and Model Theory*, §§11.4, 11.D.

From internal categoricity, we can also obtain intolerance. Informally, this says that no two $\mathrm{CLT_{int}}$-ish set concepts can diverge over any pure set-theoretic claim. Formally:[36]

Intolerance of CLT. For each second-order formula φ, whose only free variables are P and $\in$, and whose quantifiers are all restricted to P:

$$\vdash \forall P \forall \in (\mathrm{CLT}(P, \in) \to \varphi) \lor \forall P \forall \in (\mathrm{CLT}(P, \in) \to \neg\varphi)$$

The situation, then, is as with $\mathrm{PA_{int}}$. The theory $\mathrm{CLT_{int}}$ gives *internalists about set theory* a concrete response to the Skolem-Gödel Antinomy of §IV, in the specific case of the *set* concept. It explains how, using an incomplete theory, we can acquire and manifest a *set* concept which is so precise, that any purely set-theoretic claim is determinate.[37]

VIII Internalism about Model Theory

I will say more about set theory in §X. First, I want to say a bit about model theory. I dismissed modelism in §II. But my complaint against modelism is not a complaint against model theory itself. Rather, it is a complaint against a philosophical misuse of model theory. Allow me to explain.

Modelists insist on using model theory to explicate mathematical concepts. This is a mistake, as the model-theoretic arguments show. From this, Putnam correctly concluded that we must (sometimes)[38] "foreswear reference to models in [our] account of understanding" mathematical theories and concepts. But the modelist's mistake is no part of the branch of pure mathematics known as *model theory*. So, as Putnam also emphasised, we do not "have to foreswear forever the notion of a model."[39] We just need to treat the pure-mathematical *model* concept in a suitably internalist fashion.

This is quite straightforward. In common with almost every branch of mathematics, model theory is largely carried out informally: the proofs are discursive,

---

36 See Button and Walsh, *Philosophy and Model Theory*, §11.5.

37 The approach, and invocation of a technical result, is greatly indebted to Vann McGee, "How We Learn Mathematical Language," *Philosophical Review* 106, no. 1 (1997): 35–68. For more on the similarities and differences, see Button and Walsh, *Philosophy and Model Theory*, §11.A.

38 Why only sometimes? See §IX.

39 Putnam "Models and Reality," 179. See also Dummett "The Philosophical Significance of Gödel's Theorem," 191, 193.

they omit tedious steps, and so forth. But we can easily grasp 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, in principle, all of the definitions of ordinary model theory could be rewritten in austerely set-theoretic terms.

So, in what follows, let MT_{int} (for *M*odel *T*heory, *int*ernalised) be a suitable set theory to be used for model-theoretic purposes. There is no need to go into great detail about MT_{int}; I only need to explain how it relates to CLT_{int}. There are three crucial points:[40]

(1) MT_{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) MT_{int} proves CLT_{int}.
     This means that MT_{int} is internally categorical with respect to pure sets.
(3) MT_{int} proves that there are infinitely many pure sets.
     This gives MT_{int} the resources to carry out basic reasoning concerning arithmetic and hence (arithmetised) syntax.

These points ensure that MT_{int} has all the basic vocabulary and conceptual resources for developing model theory as a branch of pure mathematics. Working model-theorists will certainly want to add more axioms to the underlying set theory – I will return to this in the next section – but we need nothing more at present.

Internalists about model theory affirm MT_{int}, and insist that model theory is "officially" carried out deductively within MT_{int}. The internal categoricity and intolerance of CLT_{int} then transfers across to MT_{int}, so that any purely model-theoretic claims are determinate. As per the earlier pattern, this provides an account of how we can acquire and manifest a perfectly precise *model* concept, via a deductively-understood theory.

(At some point, of course, we might want to consider impure model-theoretic claims. For example, we might want to consider a model whose domain encompasses the cows in the field. But the internalist about model theory can deal with this straightforwardly: the specifically model-theoretic features of an impure model will be determinate, provided that there is some isomorphic model with a pure domain.)

---

40 If we want to develop an account of truth for MT_{int} itself, then we should also insist that MT_{int} is a single formula, so that we can continue to use it in the course of internal categoricity results, in the form of conditionals like $\forall P \forall \in (\text{MT}(P, \in) \to \varphi)$. For details, see Button and Walsh, *Philosophy and Model Theory*, §§12.4, 12.A.

IX Revisiting Putnam's Mathematical Internal Realism

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 Charlie has mastered arithmetic, in the form of PA_{int}, but that he knows no model theory. Nevertheless, we – who know some model theory – can pose a question: *Are any particular models of arithmetic "intended," from Charlie's perspective?*

The short answer is: *Yes: Charlie's use of PA_{int} makes certain models "intended."* But, to unpack this short answer, I will work within MT_{int}, and I will also augment the assumptions of §VIII, by assuming that MT_{int} proves second-order ZF.[41]

For any number-property, N, and any successor-function, s, let $||N, s||$ be the model (as the notion is defined in MT_{int}, of course) whose domain is the set whose members are exactly the instances of N, i.e. $\{n: N(n)\}$, and whose interpretation of the successor-symbol is the set whose members are similarly determined by s, i.e. $\{(m, n) : s(m) = n\}$. Now, the internal categoricity of PA_{int} almost immediately yields the following:

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

And we can gloss this formal result as follows:

> *MT_{int} proves that all PA_{int}-ish number concepts determine isomorphic models.*

Since internalists about model theory affirm MT_{int}, they can affirm that PA_{int} pins down a unique model (up to isomorphism). Admittedly, if our *model* concept were somehow imprecise – that is, if the *model* concept could be sharpened in different ways – then although the previous result would show that each sharpening would yield only one "model of arithmetic," different possible sharpenings

---

41 I could get away with much less than second-order ZF, but I certainly need more than just (1)–(3) of §VIII. Without some extra assumptions, I cannot prove that $\{n: N(n)\}$ exists for each property N with countably many instances, nor that $\{(m, n): s(m) = n\}$ exists for each successor-ish function s on N. So, without some augmentation, I would only be able to prove a result which we might gloss as follows: all PA_{int}-ish *number* concepts which determine a model at all determine the same model (up to isomorphism).

might allow for different "models of arithmetic." Fortunately, MT_{int}'s intolerance precludes this situation from arising: given rival sharpenings of the *model* concept, only one of them can be right.

The situation, then, is simple. Working within MT_{int}, internalists get to say that Charlie's deductive "use [of PA_{int}] already fixes the [model-theoretic] 'interpretation'."[42]

All of this has a potentially surprising consequence: internalists can (and should) agree with modelists, 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, internalists work within some model theory. And they claim to understand that model theory *deductively*, rather than *semantically*.

This observation is the key which unlocks the cryptic but beautiful closing line of Putnam's "Models and Reality." Since modelists always insist on working semantically, they embark on a futile regress, and end up treating models as "lost noumenal waifs looking for someone to name them"; this is just a poetic restatement of the lessons we learned in §II. However, by working deductively, internalists treat models as "constructions within our theory itself, [with] names from birth."[44] Saying this does not, though, involve any constructivist metaphysics; it is simply a way to summarise the central observations of this section.[45] In detail, the point is as follows: we understand our model theory deductively;[46] we define the expression "model" within that deductively-understood theory; we "construct models" by working deductively within that model theory; and we work within MT_{int} when we prove that all models of PA_{int} are isomorphic, to draw the conclusion that Charlie's deductive use of PA_{int} picks out a unique isomorphism type.

I expect the modelist to raise one last complaint against the internalist's insistence that we should understand MT_{int} deductively:

---

42 Putnam, "Models and Reality," 482.

43 Putnam, "Realism and Reason," 484, commenting on how to regard the relationship between internal realism (in general) and metaphysical realism (in general).

44 Putnam, "Models and Reality," 482. Cf. also Dummett, "The Philosophical Significance of Gödel's Theorem," 191.

45 Cf. Button, *The Limits of Realism*, 217.

46 Following Putnam, "Models and Reality," 482, I might say that "the metalanguage [i.e. the model theory itself] is completely understood." But there is a slight risk that the word "completely" might be misunderstood; treated deductively, the model theory is of course *incomplete* (on Gödelian grounds).

Working semantically, I can show that MT_{int} itself has many models if it has any. And if you insist on only ever working deductively, then you will be unable to rule out the worry that we are "trapped" in some non-standard model of MT_{in} itself. But, if you want to say that Charlie pins down the standard model by using PA_{int}, then you must rule out this worry. After all: if we are all trapped in a non-standard model of MT_{int}, then we will be right (speaking from within our non-standard model of MT_{int}) to say "all PA_{int}-ish number properties determine the same model (if they determine one at all)," but what we happen to call "the intended model of arithmetic" will be grotesque (as viewed from the outside).[47]

I can dismiss this complaint quite briskly. 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.[48] And if I can grasp *that* point, then I know that I am not "trapped" in **M**, since I just managed to quantify over the supposedly omitted elements. So: we are not "trapped" in a non-standard model of MT_{int}.

We should not infer from this, though, that we "inhabit the standard model" of MT_{int}. The same line of thought which shows that we are not "trapped" in **M** generalises to show that we do not "inhabit" any particular model of MT_{int}. Or, to drop the homely metaphors, it shows that no model of MT_{int} is "intended."[49] And if that initially sounds shocking, it really should not. Once we have abandoned modelism, there is no reason to think that a theory *needs* an "intended" model.

The overarching moral is encapsulated in a single quote from Putnam: for any theory, "either the use already fixes the 'interpretation', or nothing can."[50] But I read this as a genuine disjunction, rather than a rhetorical flourish.

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

In the case of MT_{int}, by contrast, *nothing can* fix the interpretation, for there *is* no intended interpretation (in the model-theoretic sense of "intended

---

47 Cf. Toby Meadows, "What Can a Categoricity Theorem Tell Us?" *The Review of Symbolic Logic* 6, no. 3 (2013): 539–540.

48 The modelist might object: *maybe we don't even understand (at all) the worry that we are "trapped" in **M**!* At that point, their sceptical challenge has become ineffable, and I feel we have earned the right to walk away from it. But for more on this, see Button and Walsh, *Philosophy and Model Theory*, Ch. 9, §11.6.

49 Cf. Button and Walsh, *Philosophy and Model Theory*, §11.6 on "indefinite extensibility" in this context.

50 Putnam, "Models and Reality," 482.

interpretation"). Nevertheless, our model theory is not "uninterpreted syntax." We know how to use it – deductively – and our usage manifests perfectly precise concepts. What more understanding could we want or need?[51]

X Coda: On Intolerance and Conceptual Relativity

In this paper, I have explained how internalism develops Putnam's internal realism, to provide an account of mathematical concepts which faces up to the Skolem-Gödel Antinomy. In this coda, I want to draw some speculative connections between internalism and conceptual relativism; but I relegate these remarks to a coda, precisely because they are so speculative.

In §V, I glossed the significance of PA_{int}'s intolerance result as follows: *If* Solange and Tristan share a logical language, *then* they just have to say "one of us is wrong," when one affirms Goldbach's Conjecture and the other affirms its negation. Thereafter, though, I basically acted as if the antecedent is guaranteed to hold, without further comment. So I should come clean: I cannot prove 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 ¬φ, and each could be right in their own languages.[52]

Having raised this abstract possibility, though, I should immediately point out that it is hard to see how it could actually come about. Indeed, it is not obvious that this abstract possibility is even intelligible to internalists. After all, the logical language in question is to be understood deductively rather than semantically, and we can take it for granted that Solange and Tristan accept exactly the same rules of inference. But, given this, it is hard to see what it could even *mean*, to say that they do not share a logical language.

Still, I might *just* be able to illustrate the possibility, by drawing an analogy with Putnam's discussions of mereology.[53] (To repeat: this is extremely speculative, and I am genuinely unsure what to make of it.)

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 is 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." The devil, here, will be in the details. But the specific details about mereology are not relevant here. 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 from mereology to arithmetic: *in principle*, a similar thing might happen with Solange and Tristan. If they apparently disagree, then we might (for all I know) be able to give them a suitable translation manual which smooths over the difference. And, in principle, perhaps, 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; that they are just using different words in the same situations. It is vastly harder to see what would prompt a similar thought in the arithmetical case. (I cannot think of anything, but maybe this is just lack of imagination on my part.)

It is also worth emphasising that the in-principle-possibility of reinterpreting logical vocabulary is compatible with everything I said in §VI about objects and objectivity. Tolerance concerning reinterpretation "is not a facile relativism that says 'Anything goes'."[54] It simply makes room for the in-principle-possibility that we might be free to choose between different languages, such that φ is the right thing to say in one language and ¬φ is the right thing to say in the other. Still, if Solange has fixed a language and affirms φ, and if Tristan now affirms ¬φ, then Solange must regard Tristan either as speaking falsely or as speaking a different language. Embracing this disjunction yields no sacrifice of objectivity, for it is entirely commonplace. If Tristan says "I have a pet dragon," I have the same two options – regard him either as speaking falsely or as speaking a different language – but this does not make it "up to me" whether dragons exist.

---

54 Cf. Putnam, *Reason, Truth and History*, 54.

To summarise, then, here is the more cautious statement of the significance of an intolerance theorem. For now, I am (just about, in principle) open to tolerance, when it comes to choosing a logical language. But, within a logical language, and in the presence of an intolerance theorem, divergence cannot be tolerated. Moreover, given what I just said about pet dragons, this latter kind of intolerance suffices to secure all of the objectivity – and hence all the realism – that we could ever want or need.[55]

I now want to turn from arithmetic to set theory. 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. It is less clear that this is a virtue. After all, can it really be that easy to arrive at the conclusion that the continuum hypothesis (for example) is determinate?

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. (The existence of unknowable mathematical truths is perfectly compatible with internalism.)

Second: to prove the intolerance of CLT, we need to use an impredicative principle of second-order comprehension.[56] So: maybe those who think that the continuum hypothesis is indeterminate should reject impredicativity.

Third: a few paragraphs ago, considerations about conceptual relativity led me to offer a slightly more cautious statement of the significance of an intolerance theorem. In the case of the intolerance of CLT, the more cautious statement would be as follows: 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. So, perhaps the in-principle-possibility of tolerance in *choosing a logical language* is all that is needed, for those who want to explore set-theoretic indeterminacy.

---

55 To echo Putnam, "Realism and Reason," 489 again.

56 For a proof that impredicativity is necessary, see Button and Walsh, *Philosophy and Model Theory*, §11.C.