Σ_1^0 SOUNDNESS ISN'T ENOUGH: NUMBER-THEORETIC INDETERMINACY'S UNSAVORY PHYSICAL COMMITMENTS

SHARON BERRY

Abstract. It's sometimes suggested that we can (in a sense) settle the truth-value of some statements in the language of number theory by stipulation, adopting either ϕ or $\neg\phi$ as an additional axiom. For example, in [4] and a series of recent APA presentations, Clarke-Doane suggests that any Σ_1^0 sound expansion of our current arithmetical practice would express a truth. In this paper, I'll argue that (given a certain popular assumption about the model-theoretic representability of languages like ours) we can't know ourselves to have any such freedom.

1. Introduction

It's sometimes suggested that we're (in a sense) free to stipulate either answer to certain undecidable questions in the language of number theory[1]. For example, in [4] and a series of recent APA presentations[2], Clarke-Doane suggests that mathematicians are free to adopt any Σ_1^0 sound variant on (or expansion of) our current mathematical axioms.

Call the claim that there's *some* number-theoretic sentence ϕ, such that we could safely add ϕ or $\neg\phi$ to our current number-theoretic practice, the Number-Theoretic Indeterminacy (NTI) thesis. Claiming that many different such variants on our actual number-theoretic practice are acceptable can be appealing because it promises to help solve access worries. For, crudely put, the wider the range of acceptable choices of axioms of number theory we might have adopted, the less puzzling it is that our axioms are acceptable.

[1]Here, by 'undecidable questions' I mean questions about the truth of a sentence in the language of number theory which we can neither prove nor refute using our current axioms and inference rules.

[2]In the version of this paper accepted for publication and posted 3/28/2020 some versions of these talks and [3] Clarke-Doane talks about Π_1^0 soundness rather than Σ_1^0 soundness, but this is likely a typo as all consistent theories extending Peano Arithmetic are Π_1^0 sound. In any case, my argument works against theory that all Π_1^0 soundness in exactly the same way.

In this paper, I'll present what I take to be a serious consideration against the Number-Theoretic Indeterminacy thesis (and every larger philosophy of mathematics implying it). I'll argue that accepting number-theoretic indeterminacy commits one to either making an intuitively unjustified assumption about the physical world or rejecting a certain popular assumption about the model-theoretic representability of natural languages. Thus, philosophers (who accept the relevant assumption) shouldn't take themselves to know number-theoretic indeterminacy and can't use it to help answer access worries.

In §2 I will clarify the number-theoretic indeterminacy thesis above (of which theory that all Σ^0_1 sound expansions of our current number theory are on par is a particularly attractive instance). In §3 I will explain the popular assumption about model-theoretic representability (MTR) referenced above. In §4 I'll make my main argument: that someone who knew both NTI and MTR could infer a certain contingent empirical claim that shouldn't be knowable a priori from these philosophical theses alone. Then in §5 I'll add some formal details, and in §6 I'll add some reflections and clarifications.

2. Number-theoretic Indeterminacy

2.1. **The NTI Thesis.** So let me begin by introducing the NTI thesis I mean to argue against, and explaining why it is appealing.

> **Number-Theoretic Indeterminacy Thesis**: For some number-theoretic sentence ϕ, we could equally well add either ϕ or $\neg\phi$ to our current number-theoretic axioms yielding a total mathematical practice which is 'on par' with our current one.

Note that the Number-Theoretic Indeterminacy Thesis asserts that we can add either ϕ or $\neg\phi$ to our current *overall* mathematical practice, which I take to include a priori acceptance of both first-order pure mathematical sentences (like the Peano Axioms) and some applied mathematical sentences like the following:

- All instances of the first-order induction schema in our current language including ones that involve non-mathematical vocabulary.[3]
- Certain quasi-analytic sentences about counting.

This attention to applied mathematics is important because everyone agrees that all consistent extensions of first-order Peano Arithmetic correspond to legitimate topics of mathematical investigation (many such axiom systems are studied in nonstandard number theory). The controversial claim in question is that one can harmoniously add whichever one of ϕ and $\neg\phi$ you prefer to the whole of our current mathematical practice. We can only hope to use Number-theoretic Indeterminacy to reduce access worries if we can show that variants on our actual mathematical reasoning methods — which include a priori acceptance of the impure mathematical sentences above- would have been equally truth good/truth conducive.

I take the claim that adding either ϕ or $\neg\phi$ as an axiom would yield a mathematical practice on par with our current one to imply at least that if someone made one of these choices then: the axiom they added would express a truth (in their language), while all our current mathematical axioms and the pure number-theoretic sentences mentioned above would continue to express truths (in their language).

2.2. **Two Types of NTI Theories.** Now why accept the above number-theoretic indeterminacy thesis? As noted above, accepting NTI seems to help with access worries by allowing that a wider range of variant mathematical theories would express truths. Perhaps it can also be motivated by Putnamian model-theoretic arguments suggesting that we can't pin down a unique intended natural number structure[9]. However, it will be helpful to distinguish two kinds of number-theoretic indeterminist views.

---

[3]One such instance of the induction schema is, "If some book has 0 pages, and for each number n, if some book has n pages then some book has $n+1$ pages, then for every number there's some book which has that number of pages."
One might worry that this is too strong. For some philosophers to respond to Sorites problems by saying that induction can fail for cases involving vague predicates, e.g., 'if there's a no heap consisting of 0 and for all n if there's no heap consisting of n grains of sand then there's no heap consisting of $n+1$ grains of sand don't...'. However, even if one takes this view it causes no trouble for me, as my argument can be easily modified to use only non-vague physical predicates and relations.

One type of number-theoretic indeterminist view says mathematical truth shouldn't outrun proof, and so takes *every* number-theoretic sentence ϕ which can't be proved or refuted[4] to be indeterminate in a way that implies the mathematical practices which stipulate either its truth or falsehood are on par with each other (and our own mathematical practice). It holds that all *syntactically consistent* variants on our current number-theoretic practice would express truths[5]. I will, following [6], call this view Extreme Anti-Objectivism.

Famously, Extreme Anti-Objectivism faces a worry involving the apparent connection between number theory and facts about consistency[6]. For, the above proposal appeals to the idea that there are definite facts about consistency. But consistent variants on our current arithmetical practice can (seemingly) take different views about provability and consistency through implying different arithmetical Con() sentences[7]. And it can seem odd to be a realist about consistency while saying that variants on our mathematical practice which, seemingly, get consistency facts wrong (e.g., claim that a consistent theory is actually inconsistent) would be on par with our current practice[8].

Another type of view implying NTI is what I'll call 'moderate anti-objectivism'. Philosophers of this stripe avoid the problem above by saying certain incompatible ways of extending our current number-theoretic practice are legitimate, but only ones *which don't get consistency wrong* (in the sense above). The simplest example of such a view is Justin Clarke-Doane's suggestion (in [4] and a series of APA

[4]That is, every claim which can't be proved or refuted from our current axioms.
[5]Henceforth by 'consistent' I will mean syntactically consistent, rather than any richer notion of coherence like that invoked in [2], or having a model which treats second-order quantifiers standardly.
[6]For example see Hartry Field's discussion of 'extreme mathematical anti-objectivism' in [6].
[7]By Gödel, there are consistent extensions of Peano Arithmetic (or any recursively axiomatizable extension thereof) which disagree on the truth of $\mathrm{CON}(T + \phi)$ where T is our current number-theoretic practice and ϕ is a sentence in the language of number theory. Specifically, we're inclined to accept all instances of the schema 'Con(S) iff S syntactically consistent' for a suitably concrete descriptions of some axiom system S, so it seems good interpretations of number theory should make this claim true.
[8]See chapter 3 of [7] for a careful development of this argument. I personally take no stand on how successful this argument is, as my main arguments attempts to show that no theory implying the NTI thesis should be accepted. I merely mention it as a common motivation for favoring weaker number-theoretic indeterminacy views implying NTI.

presentations[9]). that all (and only) Σ^0_1 sound extensions of number theory (those which don't prove any false sentences in the language of arithmetic with only a single unbounded existential quantifier) are acceptable. Saying this ensures that only expansions of our number-theoretic practice which don't get consistency facts wrong are acceptable[10].

Mere Σ^0_1 soundness and its ilk can seem to provide a principled and attractive middle ground between realism and antirealism concerning the numbers. On one hand, we avoid the problem for extreme antiobjectivists above by saying that all claims of the form "no natural number codes a proof of '0=1' in formal system S" have definite right answers; (just as one would hope) they're determinately true if the formal system in question is consistent and determinately false otherwise. On the other hand, one reduces the strength of access worries (compared to more truth-value realist views) by saying getting mathematics right doesn't require recognizing facts about the coherence of theories logically powerful enough to pin down right answers to all questions in number theory (like Second-Order Peano Arithmetic) or truth in some favored platonic model of the natural numbers[11][12].

However, I'll argue that a certain popular assumption about model-theoretic representability is difficult to reconcile with the Number-theoretic Indeterminacy thesis — and hence with even modest antiobjectivist views, like the mere Σ^0_1 soundness thesis.

[9]Clarke-Doane now rejects this view in favor of taking there to be definite right answers to all questions in the language of arithmetic[3].

[10]This is true because, as noted above, syntactic consistency claims are intuitively true if and only if certain con() sentences in the language of arithmetic are true, and each of these is explicitly a Σ^0_1 sentence.

[11]For example, if you take our determinate reference to the natural numbers to be secured by employing full second-order quantification, the same resources will also let us pin down the width of the hierarchy of sets and hence the truth-value of CH.

[12]A further appealing feature of this weakening is that any theory which avoids proving false Con sentences will automatically avoid the kind of counterargument by an imaginable physical experiment involving infinitary computers discussed in [2]. For these hypothetical experiments with (simple, less controversially imaginable) hypercomputers showed the falsehood of some Σ^0_1 sentence independent of our current arithmetic.

3. Model-Theoretic Representability Expectations

I take it that many philosophers accept something like the following idea about the philosophical illumination provided by set-theoretic models and Tarski's definition of truth.

We can illuminatingly think of truth-values for our utterances in English and other natural languages as being built up from ground level facts about the references of names and extensions of predicates etc. in accordance with Tarski's recursive definition of truth. Thus (if we temporarily bracket certain issues about the size of the universe, vagueness etc. discussed in A and 6.1) we'd expect there to be a set-theoretic model which assigns references and extensions to terms in our language correctly in such a way that applying the Tarskian recursive definition yields correct truth-values for all our sentences.

Furthermore, something similar goes for speakers who mildly change the meaning of certain terms by adopting an additional mathematical axiom. Specifically, adding such axioms can't and won't change the overall Tarskian structure of their/our language. It also, plausibly, won't change the meaning of certain concrete/physical vocabulary like 'cat' and 'bites', whose meaning is fixed by observation procedures or deference to natural kinds. It can only sharpen or change the meaning of mathematical and more abstract terms so as to (if possible) ensure the truth of the resulting of one's resulting total collection of mathematical axioms[13].

Thus, (continuing to ignore concerns about size etc.) there should also be an **adequate set-theoretic model for a person's language**, in the following sense.

- This model interprets physical terms like 'cat' and 'bites' standardly.
- Applying Tarskian definitions to this model yields correct truth-values for all the speaker's sentences (or at least all the speaker's sentences that are determinately true or false).

[13]I take it that if we added axioms in a way that made our total axiom system inconsistent, or allowed one to derive that the total universe contained only three things.

Note that the concept of being an adequate modeling a person's language (in the sense defined above) is very different from the concept of satisfying a person's best theory of the world. An adequate model has to get the *actual truth-values* for certain sentences in a person's language right. But it doesn't have to make all their *beliefs* come out true. Indeed typically an adequate model of a speaker's language can't make all their beliefs come out true, because most people believe at least one thing that's determinately false.

Some have tried to motivate versions of the model-theoretic interpretability (MTR) assumption above by a kind of inference to the best explanation. For example, it has been suggested that Tarski's theory provides the best (or only known decent) candidate for an explanation for the finite learnability of human languages[5] or the complex pattern of truth conditions English sentences take on a general theory of truth[14]. But I take the above MTR principle to be widely accepted, whether or not these projects succeed.

4. Main Argument

If number-theoretic indeterminacy is true, two speakers Alice and Bob can make syntactically incompatible but acceptable choices for which axioms of number theory to add. In particular, the NTI theorist thinks the following is possible.

> **Acceptable Divergent Axiom Choice** We start by accepting Peano Arithmetic together with certain conceptually core sentences

[14] In [1] Benacerraf writes, "any theory of mathematical truth [should] be in conformity with a general theory of truth-a theory of truth theories ...which certifies that the property of sentences that the account calls 'truth' is indeed truth. This... can be done only on the basis of some general theory for at least the language as a whole (I assume that we skirt paradoxes in some suitable fashion)... the semantical apparatus of mathematics [should] be seen as part and parcel of that of the natural language in which it is done, and thus whatever semantical account we are inclined to give of names or, more generally, of singular terms, predicates, and quantifiers in the mother tongue [should] include those parts of the mother tongue which we classify as mathematics. But regardless of the success of these projects, I take MTR to be widely accepted.
I take it that we have only one such account: Tarski's, and that its essential feature is to define truth in terms of reference (or satisfaction) on the basis of a particular kind of syntactico-semantical analysis of the language, and thus that any putative analysis of mathematical truth must be an analysis of a concept which is a truth concept at least in Tarski's sense. "

> about applied mathematics[15]. Alice and Bob do the same. But, for some number-theoretic sentence ϕ independent of our current number-theoretic practice, Alice adds[16] ϕ to her theory of arithmetic and Bob adds $\neg\phi$ to his theory. Afterward, all the sentences Alice now takes to be axioms or conceptually central truths about the numbers (as per §2) express truths when spoken by her. The same goes for Bob.

Note that, by the completeness theorem there are (i.e., our current set theory includes) models of the 'natural numbers' which make Alice's axioms come out true and other models which make Bob's axioms come out true. However, at least one of those models must interpret Alice's term 'number' non-standardly, adding 'infinite' natural numbers which are larger than all finite natural numbers (while being cleverly structured to keep all instances of the first-order induction schema in the language of number true). Thus, at least one of Alice and Bob can only be model-theoretically interpreted (from the point of view of our set theory) by taking them to be talking about something which isn't isomorphic to the natural numbers[17].

Without loss of generality, assume that Alice is the one whose pure mathematical axioms require us to interpret her non-standardly. This fact alone doesn't prevent Alice from being model-theoretically representable. For recall that the model-theoretic representability assumption doesn't require that we interpret her mathematical expressions like 'number' as referring to the same objects as ours do. It merely requires interpreting physical vocabulary like 'Emperor' and 'ruled after' standardly.

---

[15] Note that the standard induction schema in Peano Arithmetic automatically lets us perform induction on any predicate definable in our language, even predicates which make use of non-number-theoretic vocabulary.

[16] This talk of adding is merely a verbal flourish. It is enough to imagine Alice simply accepting our practice as well as the truth of ϕ.

[17] Note that neither Alice and Bob accept any false Δ^0_1 sentences. That is, both their theories are Δ^0_1 sound.

However, I'll now show that we can, by otherwise uncontroversial reasoning, show that if Alice's language can be attractively model-theoretically represented (as follows from Acceptable Divergent axiom choice together with the model-theoretic representability assumption MTR in §3), then a certain empirical hypothesis which we intuitively shouldn't be able to decide by a priori philosophical reflection alone is false. That is, working within our current set theory we can prove the following conditional, 'If there are an ω sequence of emperors then there are not attractive model-theoretic representations of both Alice and Bob'. Note that I'm showing there can't even be *separate* models that attractively interpret both Alice and Bob's speech; I'm not assuming that Alice and Bob can refer to each others' number concepts[18].

Now I will introduce the epistemic possibility which makes it difficult to adequately model-theoretically represent Alice[19].

> **ω-sequence of Emperors**: The Roman Empire will be reestablished, and time will turn out to be infinite and the laws of entropy an illusion etc. so that Virgil's dream of imperium sine fine literally comes true. Specifically, there is (what our current set theory with ur-elements classifies as) an ω sequence of emperors.

To see why Alice can't be adequately interpreted in this case, first, note that if we interpret 'is the n-th emperor' standardly[20] then we're forced to interpret Alice's '... is the ...-th emperor' as relating the emperors to exactly the standard initial segment of this nonstandard model. For it seems this relation should pair Augustus with the first of Alice's 'numbers', Tiberius with the least 'number' after that, Caligula

[18]Thus I won't make the assumption that Alice can somehow refer to Bob's 'numbers' and use them to define an initial ω sequence and thus contradict the induction axiom as per [8]. For the NTI theorist isn't committed to the possibility of a unified language which (so to speak) quantifies over both Alice's and Bob's numbers (and makes each of their axioms come out true about their own copy of the natural numbers).

[19]Note that, as will be discussed in §6 my choice to use the terms 'emperor' and the relation 'ruled after' here is made solely for vividness, and any physically definable relations would do. Those who think induction can fail when vague non-mathematical properties are involved might replace these physical terms with ones they think are guaranteed to apply determinately.

[20]That is if we identify the initial segment of Alice's 'numbers' with our numbers then the extension of the relation 'is the n-th emperor' would have the same extension for Alice as it does for us.

with the least after that etc. until we have mapped the emperors to the standard initial segment of the Alice's 'numbers', at which point we have no emperors left to pair with the remaining 'nonstandard numbers' in our nonstandard model. In this way, we map onto all and only the standard initial segment of Alice's numbers (note that an adequate interpretation can't change the extension of emperor). But then we must interpret the property $Q(n)$ "there is an n-th emperor" as applying to 1 and the successor of every 'number' it applies to, but not to all numbers. So we must interpret the following claim as coming out false.[21]

> **Informal Emperor Induction**: If there's a first Roman Emperor and, for every number n, if there's an n-th emperor then there's an n+1st emperor then for all numbers n there's an nth emperor.

However, on our description of the NTI's Acceptable Divergent Axiom Choice scenario this instance of the induction schema expresses a truth in Alice's language. Thus, there can be no model-theoretic interpretation which gets Alice's truth conditions right

OK but what if we don't interpret the two-place relation'is the n-th emperor' standardly? Insofar as this relation is a somewhat mathematical one, it seems reasonable that adopting a new axiom of number theory could change the meaning of this mathematical sounding relation. However, allowing this kind of wiggle room turns out not to help. For it turns out that merely ensuring the truth of certain sentences which seem to be (or follow from) conceptually core truths about counting and the numbers forces us to interpret 'is the nth emperor' standardly in the sense above, i.e., as relating the emperors to exactly the standard initial segment of (our model for) Alice's 'numbers'. I lay out this argument in detail in §5.

So the argument above goes through just the same. Thus if there are an ω sequence of emperors, any attempt to adequately model-theoretically represent Alice must fail.

[21]The last two points follow from the clause concerning Tarskian recursive definitions of truth in our definition of model-theoretic representability

So it follows from the model-theoretic interpretability assumption (and the fact that we can't know there won't be an ω sequence of Roman Emperors) that Number-theoretic Indeterminism is an unattractive view. More specifically, anyone who takes themselves to know the truth of the NTI thesis is committed to the possibility of superficially similar but deeply model-theoretically unrepresentable languages.

5. Details of Non-Modelability

To articulate the above argument that one can't adequately model-theoretically represent Alice (given the assumptions about Alice and the Roman Emperors made in S4) more rigorously, I'll need to formalize a relevant fragment of Alice's language and counting talk.

Consider our use of the natural numbers to talk about temporally ordered objects and events such as 'the 4th U.S. President', 'the 37th successful rickrolling', or 'the 3rd Roman emperor'. Now I take all the sentences below to be (or be derivable from) conceptually central truths - which any acceptable interpretation of Alice must make true - regarding counting (temporal) sequences of events using the natural numbers.

COUNTING RULES

- An object x is the 0th emperor, i.e., $countemp(0, x)$ iff x is a emperor and all other emperors happen after x[22].
- If x is the nth emperor, then y is the $n + 1$th emperor iff y occurs after x and no other emperor occurs between x and y [23].
- Only emperors can be the nth emperor[24].
- No two distinct numbers correspond to the same emperor[25].

[22]That is, $(\forall x)[countemp(0, x) \leftrightarrow emperor(x) \wedge (\forall y)(countemp(y) \rightarrow before(x, y) \vee x = y))]$. Here I am departing from common practice in counting up from 0 rather than 1, for the sake of simplicity. Also by writing this formula with 0, I abbreviate the corresponding claim about the unique object satisfying some definite description of 0 in terms of $\mathbb{N}.S$, e.g., 'the unique object that has a successor but isn't a successor.' .

[23]That is,

$$(\forall n)(\forall x)(\forall y)(\mathbb{N}(n) \wedge countemp(n, x) \rightarrow [(countemp(S(n), y) \leftrightarrow$$
$$emperor(y) \wedge before(x, y) \wedge (\forall z)\neg\, (emperor(z) \wedge before(x, z) \wedge before(z, y))]$$

[24]That is $(\forall x)[(\exists n)countemp(n, x)) \rightarrow emperor(x)]$.

[25]That is $(\forall x)(\forall n)(\forall m)[emperor(n, x) \wedge emperor(m, x) \rightarrow m = n]$

One can formulate these sentences in first-order logic as indicated in the footnote above (where $emperor(x)$ denotes 'x is a Roman emperor', $countemp(n, x)$ denotes 'x is the n-th emperor' and $before(x, y)$ denotes that the emperor x ruled before emperor y). And we can use the same vocabulary to formalize the predicate $Q(n)$ "there is an n-th emperor" above $Q(n) \stackrel{\text{def}}{=} (\exists x)(countemp(n, x))$. So the instance of the induction schema above becomes the following.

So our Alice will also accept the following formal sentence as a conceptually central mathematical truth.

EMPEROR INDUCTION $Q(0) \wedge (\forall n)\,[Q(n) \rightarrow Q(S(n))] \rightarrow (\forall n)[\mathbb{N}(n) \rightarrow Q(n)]$

Now to see why we can't simultaneously interpret Alice's number talk as referring to a nonstandard model and her terms 'emperor()' and 'ruled before()' as referring standardly (in the scenario under consideration) while making both her sentences COUNTING RULES and EMPEROR INDUCTION come out true, consider our predicament when choosing an extension for her relation '*countemp*()' (i.e., '..is the ...th Roman Emperor').

We are forced to interpret '*countemp*()' as relating the emperors to exactly the standard initial segment of Alice's 'numbers' for the following reason. We can only satisfy the first element of COUNTING RULES (while interpreting 'emperor()' and 'before()' standardly), if we interpret '*countemp*()' as relating 0 to the temporally first Roman Emperor (Augustus). And, given that we've done this, we can only satisfy the second conjunct of COUNTING RULES if we interpret '*countemp*()' as relating 1 to the second Roman emperor (Tiberius), and so on for all the objects in the standard initial segment of the nonstandard model. But this 'uses up' all the Roman emperors (and recall that we're taking Alice's term 'emperor()' to apply to all and only the Roman emperors). Since, by the last conjunct of COUNTING RULES, no emperor can be counted again (i.e., associated with another putative natural 'number') it follows that all and only the 'true' (i.e., standard) natural numbers are paired with emperors.

So we must make '$Q(0)$' and 'Whenever Q applies to some number n it also applies to $Q(n+1)$' true. But as we take 'the numbers' to refer to a larger structure than this standard initial segment, our interpretation also makes '$\forall n Q(n)$' false. So our interpretation renders EMPEROR INDUCTION false.

5.1. **Summary From A Determinist Point of View.** One can think of what is going on here (from a determinist point of view) as follows. Initially, if seems like we can reconcile NTI with model-theoretic intepretability by simply modeling speakers who stipulate different consistent extensions of the Peano Axioms as speaking about nonstandard models of the natural numbers. The fact that some of these models must not be isomorphic to our current model of the numbers seems like a small cost (if any). For it's plausible that adding new axioms of number theory could change the kind of structure your number talk would have to be about.

However, it's implausible that adding new axioms of number theory could change how terms like 'emperor' and 'ruled before' would apply. And this creates a problem for interpreting Alice using a nonstandard model.

For the nonstandard models are cooked up to satisfy all our first-order pure mathematical axioms, including induction for purely mathematical predicates, but they don't satisfy real second-order induction. And appealing to physical stuff can let us reveal that, by picking out portions of mathematical structures we couldn't pick out using pure mathematical language alone[26]. There are some ways that we consider it epistemically possible for 'emperor' and 'successor' to apply which would let Alice define subsets she couldn't with pure mathematical vocabulary alone. More specifically, if our numbers are an initial segment of hers and the emperors have the structure of our numbers, her predicate $\phi(n)$ 'There is an emperor who is the n-th emperor' applies to a successor closed initial segment of Alice's natural numbers — hence induction fails for this predicate.

[26]For another example of how the physical world can let us secure reference we might not otherwise be able to, consider reference to physical constants, i.e., constants in the laws of physical nature, could let us refer to real numbers which we can't provide definite descriptions for in our countable pure mathematical language.

6. Reflections and Clarifications

6.1. **Clarifications.**
Next let me clarify some points about the shape of my argument, the assumptions it requires.

First, does this argument question-beggingly presume determinate reference to the natural numbers in the metalanguage we use to reason about Alice and Bob? After all, the argument in §4 and §5 asks us to assume in the metalanguage that the emperors form an ω-sequence. So doesn't that implicitly assume determinate reference to the natural numbers? If so, then the above argument would indeed be question-begging.

However, my argument requires no such determinacy assumption, which we can see as follows. Whatever your views on reference, considering the line of reasoning in §4 and §5 reveals the following fact. If someone one knew both MTR and NTI, they could derive by standard applied set-theoretic reasoning[27] that, 'The emperors do not form an ω sequence.' And the latter claim is (intuitively and in terms of ordinary scientific practice for reasoning) a contingent matter which we shouldn't be able to learn by a priori reflection alone[28]. Note that, under many standard views about indeterminacy, claims you can prove from determinately true claims are still determinately true even if the references of some of the concepts are indeterminate. Indeed, if you believe that the claim, "The emperors are isomorphic to an ω sequence" is indeterminate (in at least some epistemically possible scenarios) then you can't allow that it's provable from determinately true claims.

Now my target (the advocate of both MTR and NTI) could, in principle, bite this bullet, accepting that we have a priori knowledge of the seemingly contingent

[27]More strictly speaking, it demonstrates that someone who knew these things could derive the relevant conclusion using only standard set-theoretic reasoning and the inference from MTR and NTI to the Two Models claim defined below.

[28]Note, because the indeterminist doesn't think we pick out a single intended structure with our number talk, they might say it's epistemically possible that 'The emperors form an ω sequence.' is *indeterminate* rather than epistemically possible that this claim is *true* on the following grounds. It's epistemically possible that the emperors under 'ruled after' are isomorphic to *one* candidate natural number structure. But in this case the emperors won't be isomorphic to any of the other non-isomorphic candidate natural number structures. However, taking this position makes no difference to our argument. For one still shouldn't be to prove 'The emperors *aren't* isomorphic to the numbers' a priori from philosophical premises that are (presumably) determinately true.

claim that the emperors don't form an ω sequence. But I think this is a sufficiently unattractive consequence to be a genuine weakness of that view. Also, note that biting this bullet would yield something analogous to a Quinean indispensability problem. For 'The emperors aren't isomorphic to the natural numbers' abbreviates a perfectly ordinary applied mathematical sentence using no special philosophical or logical vocabulary[29] which we'd normally say can't be assumed a priori. So they'd have to revise applied mathematical practice to allow for a priori knowledge of such claims. And it's not clear how to do that.

Second, one might worry that the assumption of MTR is obviously incompatible with NTI, as it requires to be determinately true or false (as models can only represent sentences as true or false). Yet this conflicts with the intuition behind NTI which pushes us to say that (a version of) NTI should hold for Alice and Bob's number-theoretic practice as well as our own. Thus, MTR is obviously incompatible with the kind of argument the proponent of NTI wants to make.

In response, I want to point out two things. For one thing, there's no immediate conflict between MTR and NTI. The indeterminist can take MTR to express a truth on all acceptable sharpenings of their set-theoretic talk just (as they likely think ZFC does). The fact that Alice and Bob's number talk is indeterminate doesn't necessarily cause a problem since the same indeterminacy arises for our interpreter.

For another thing, the argument of sections 4 and 5 works just the same if we replace the premise MTR with a weaker claim MTR* that (as per the supervaluationist idea above) only requires the existence of models for a speakers' language that give sentences that are determinately true or false in that language the right truth-value (and get the extension of physical vocabulary right). Indeed, it works using only the following premise (which follows from NTI and MTR* by considering any pair of models M_a and M_b providing an acceptable sharpening of Alice's and Bob's notions respectively)[30].

[29]'There's no (set coding a) 1-1 onto function f from the natural numbers to the emperors such that $n < m$ iff $f(n)$ ruled before $f(m)$.

[30]Note that taking an NTI theorist (call her Claire) to know Two Models and expresses a determinate truth by it, doesn't require thinking that *she* has determinate reference to sets in the

Two Models There's a number-theoretic sentence ϕ such that (in set theory with ur-elements) there's a model which gets the extensions of 'emperor' and 'ruled after' right and makes PA, EMPEROR INDUCTION, COUNTING RULES and ϕ come out true, while getting the extensions of physical vocabulary right. Similarly, there's a model which does exactly the same except that it makes $\neg\phi$ rather than ϕ come out true.

For the argument of sections 4 and 5 shows that anyone who accepted Two Models could use standard set-theoretic reasoning to derive a conclusion (the emperors aren't isomorphic to the numbers) which we shouldn't be able to derive a priori. So it doesn't just show the NTI theorist shouldn't take themselves to know MTR, but also that they shouldn't take themselves to known MTR* or even Two Models.

Second note that, for the purposes of my argument it doesn't matter whether anyone can learn any contingent truths about how many Roman emperors there are. It also doesn't matter whether Alice and Bob accept set-theoretic notions and could give a version of the argument in §4 and 5 above.

For the trouble isn't about what Alice and Bob could discover, but rather about what the indeterminist herself can prove by thinking about Alice and Bob's languages. If my target number-theoretic indeterminist really knew both MTR and NTI, then *she* could (by uncontroversial set-theoretic reasoning) prove a contingent claim about the structure of the Roman empire which shouldn't be derivable from a priori philosophical reflections.

Third, I playfully formulated my argument in terms of claims about how the Roman emperors could be related under 'ruled after' for simplicity, picturesqueness and a Frege reference. However, we could make exactly the same argument using

---

metalanguage. For we can say that there are many acceptable precicifications of Claire's talk all of which render Two Models determinately true (making Two Models and all consequences of it determinately true).

Also note that to a supervaluationist who accepts this, needn't think all (or any) of precificications of Claire's set talk will capture the true range of possible precifications of Alice's talk. It's enough that each acceptable interpretation of Claire's set-theoretic talk includes enough models to witness the indeterminacy of each sentence in Alice's language.

any other physically definable predicate P and relation $>_p$ instead (e.g., sequence of temporal points invoked by Field's proposal discussed below).

6.2. **Contrast With Field.** Finally, my argument for the conditional 'If there's an ω sequence of emperors Alice's set talk can't be adequately model-theoretically represented' takes inspiration from Hartry Field's work in [7], but attempts to remove certain unacceptable assumptions needed for Field's story.

Field, in effect, points out that if time has a certain plausible structure (and determinate reference for physical vocabulary can be assumed) we could 'define' the intended natural number structure by saying that the numbers are isomorphic to a certain sequence of points in time (describable using no mathematical vocabulary). And he tentatively suggests this could explain our ability to secure determinate intended meanings for expressions like 'finitely many' and (up to isomorphism, ignoring concerns arising from Field's nominalism) 'natural number'.

However, contrary to the latter hope, I claim Field's story can't explain how our natural number talk has determinate reference (even in the case of English speakers who happen to believe that the natural numbers are isomorphic to Field's sequence of temporal points). Call the sentence asserting the latter isomorphism ISO. The problem is that even those who take themselves to know ISO accept it as a contingent physical hypothesis, not a definitional truth constraining all acceptable interpretations of our expressions 'natural number' (or 'finite') [31].

More specifically, Field's story only succeeds in explaining how determinate reference for our 'number', if all acceptable interpretations of us must make ISO come out true. But the latter is not a plausible constraint on acceptable interpretation of people who believe ISO. For accepting it yields the repugnant conclusion that if we'd lived in a (experientially indistinguishable) world where time had a different structure, so that the temporal points picked out by Field's description formed

[31]Note that my model-theoretic representability assumption (MTR) above only requires that Alice's sufficiently conceptually central beliefs be interpreted to come out true, not all her empirical hypotheses

a *nonstandard* model of number theory, we'd have meant this *nonstandard model* when we talked about 'the numbers'.

My story fixes this problem by replacing the false assumption that acceptable (non-trivial) interpretations of our number talk must make our belief in ISO come out true with the more plausible assumption that they must make (necessary, a priori, quasi-analytic) sentences like EMPEROR INDUCTION and COUNTING RULES come out true[32].

Here's another way of putting the point. Just because one could stipulatively fix determinate reference for my word 'number' by stipulating that the 'natural numbers' are to have the same structure sequence of temporal points, doesn't imply that currently (not having made such a stipulation) my reference is currently determinate, or I'm that not free to adopt any Σ_1^0 sound extension of PA as an axiom[33].

7. Conclusion

In this paper, I've argued that accepting any form of number-theoretic indeterminacy has a grave cost — or at least an important consequence. Taking ourselves to know that even one number-theoretic sentence can be safely settled by stipulation turns out to conflict with very common assumptions about the model representability of natural languages.

---

[32]Because the latter are necessary truths, there's no bar to saying that intrinsic duplicates of us living in worlds where time has a weird structure (so the emperors/temporal points form a nonstandard model of number theory) can be interpreted as meaning the same thing as we do by 'number' and having their sentences EMPEROR INDUCTION and COUNTING RULES come out true.

[33]Even if the number-theoretic indeterminist takes themselves to, in some sense, know 'such-and-such temporal points are isomorphic to the numbers' it doesn't follow that they're committed to this sentence continuing to express a truth in Alice's language (so that all acceptable model-theoretic representations of her must get the truth-value for this sentence right). It wouldn't be surprising if sharpening the meaning of your mathematical terms by adopting new number-theoretic axioms could change the truth-value of some scientific theories involving quantification over the numbers. In contrast, the NTI theorist is committed to thinking instances of the induction schema and quasi-analytic truths like COUNTING RULES and EMPEROR INDUCTION remain true in Alice's language.

For, from common applied mathematical assumptions like the axioms for ZFC with ur-elements[34], we can prove the following conditional, 'If there is an ω sequence of emperors then there are not adequate model-theoretic representations for the languages that would be spoken by two people, Alice and Bob, who have successfully stipulated opposite truth-values for some number-theoretic sentence ϕ (as per NTI)'. Thus any philosopher who genuinely knew both NTI and MTR (and the standard set-theoretic axioms) could derive 'The emperors don't form an ω sequence'[35]. But the latter is intuitively a contingent claim that shouldn't be knowable by a priori reflection alone.

Let me end with some quick caveats. The argument in 4 isn't intended to suggest (or assume) that anyone could recognize that there's an ω sequence of emperors[36] if there happens to be one. Nor does it suggest any way of learning *which* one of two people who've adopted syntactically inconsistent number-theoretic axioms got things right. It also doesn't suffice to block Putnamian model-theoretic skepticism, even taking determinate reference for physical vocabulary is granted. For we plausibly don't know whether there is an ω sequence of emperors (or the like[37]). So we plausibly don't know that any physical objects form the kind of short infinite linear orders needed to block nonstandard interpretations of our number theory (though I'm tempted to say we'd have to be deeply unlucky if this is not the case[38]).

In this paper I've only tried show that philosophers who accept certain common assumptions (about model-theoretic representability and determinate physical reference) can't take themselves to *know* that some number-theoretic question can be

[34]Strictly speaking not all of these axioms are needed to reconstruct the argument above.
[35]i.e., 'It's not the case that there's a 1-1 onto function f from the natural numbers to the emperors such that $n < m$ iff $f(n)$ ruled before $f(m)$'.
[36]It also doesn't imply that they could determinately refer to the structure which the emperors under successor happen to have, in any other way than as 'whatever structure the total plurality of Roman emperors happen to form when considered under the relation 'ruled before'.
[37]That is, we don't that some physical objects form an ω sequence when considered under some physically definable relation $<_p$.
[38]Also see end of REDACTED for some ideas about how combining Davidsonian theses about charitable interpretation with (a variant on) the above argument involving COUNTING RULES above might do the trick.

settled by stipulation, i.e., that there's even one number-theoretic sentence ϕ such that we could safely add either ϕ or $\neg\phi$ to our current mathematical axioms.]

Appendix A. Weakening Model Theoretic Representability

Aside from the issues addressed in §6.1 there are two reasons we might expect model-theoretic interpretability to fail. However, weakening MTR to allow for these failures makes no difference to my argument.

First, (as mentioned in the Benacerraf quote above) there are concerns about representing truth conditions in any language containing a general truth predicates. One might also worry about fitting claims involving primitive physical or metaphysical possibility operators into the picture above. We could handle this problem by restricting our adequate modelability requirement to the collection of sentences in speaker's languages which (is quantifier restricted to and) talks only about numbers and non-modal properties of physical objects.

Second, one might think the modelability requirement can fail when limitations of size are all that prevent modeling (e.g. if adequately model theoretically representing your own language required assigning 'set' a proper class sized exstension). However, weakening the model-theoretic interpertability assumption in this way does nothing to help the NTI theorist. For our problem will be that all set-theoretic interpretations make a certain speaker's (Alice's) 'number' structure too large, not that our models of her language didn't have a large enough domain of objects to work with. For note that temporarily adopting an expanded language which adds some layers of proper classes atop our hierarchy of sets helps with the problem above, but wouldn't block the argument above.

References

[1] Paul Benacerraf. Mathematical Truth. *Journal of Philosophy*, 70:661–80, 1973.

[2] Sharon Berry. Malament-Hogarth Machines and Tait's Axiomatic Conception of Mathematics. *Erkenntnis*, 79(4):893–907, 2014.

[3] Justin Clarke-Doane. *Morality and Mathematics*. Oxford University Press, 2020.

[4] Justin Clarke-Doane. Set-Theoretic Pluralism and the Benacerraf Problem. *Philosophical Studies*, 177(7):2013–2030, 2020.

[5] Donald Davidson. Truth and Meaning. *Synthese*, 17(1):304–323, 1967.

[6] Hartry Field. *Realism, Mathematics & Modality*. Blackwell Oxford, 1989.

[7] Hartry Field. *Truth and the Absence of Fact*. Oxford University Press, 2001.

[8] Vann McGee. How We Learn Mathematical Language. *Philosophical Review*, 106(1):35–68, 1997.

[9] Hillary Putnam. Models and Reality. In *Realism and Reason*. Cambridge University Press, 1983.