Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

In Defense of the Implicit Commitment Thesis

ETHAN BRAUER
Lingnan University

The implicit commitment thesis is the claim that believing in a mathematical theory *S* carries an implicit commitment to further sentences not deductively entailed by the theory, such as the consistency sentence *Con*(*S*). I provide a new argument for this thesis based on the notion of mathematical certainty. I also reply to a recent argument by Walter Dean against the implicit commitment thesis, showing that my formulation of the thesis avoids the difficulties he raises.

1. Introduction

Gödel's famous incompleteness theorems show that for any consistent theory S which is recursively axiomatizable and can interpret Robinson arithmetic, there will be true arithmetical sentences that cannot be proved by that theory. One such sentence will be *Con*(S), which expresses the consistency of S. But it would be odd to believe in S without believing in S's consistency. As Halbach (2011: 322) puts it: "Accepting a theory without believing in its consistency strikes many logicians at least as odd if not incoherent. If one endorses a theory, so one might argue, one should also take it to be sound," and thus consistent. On this basis, many authors have accepted what Dean (2014) calls the *Implicit Commitment Thesis* (ICT), which he formulates as follows. (I will offer a slightly refined version of the thesis in §2).

> Anyone who accepts the axioms of a mathematical theory S is thereby also committed to accepting various additional statements Δ which are expressible in the language of S but which are formally independent of its axioms. (2014: 32)

Contact: Ethan Brauer <eebrauer@gmail.com>

https://doi.org/10.3998/ergo.3114

Dean gives a more critical treatment of ICT, however, arguing that the viability of certain foundational views of mathematics undermines ICT. His argument is that these foundational views justify theories S that are, as he calls them, *epistemically stable*, meaning that S is justified by a foundational view which justifies no stronger theory. If Dean's argument is successful, this would undermine ICT by showing that there are theories S which can be justifiably accepted without incurring a commitment to any independent statement Δ. For otherwise, one would be justified in accepting $\mathsf{S}+\Delta$, contradicting the epistemic stability of S.

My goal is to defend ICT. In §2 I will provide a positive defense of ICT by arguing that when you have mathematical certainty in a theory S you are also in a position to be mathematically certain of its local reflection principles. While the basic idea of the ICT is a familiar one, the arguments offered for this claim are often imprecise and impressionistic. I spell out in precise detail a novel argument, paying particular attention to exactly what assumptions about S are required. In particular, I emphasize that there must be a description of S that you can recognize as a set of statements you accept. I also discuss the extent to which my arguments might apply to theories that do not enjoy mathematical certainty. In §3 I then critically examine Dean's two case studies of foundational views that allegedly justify epistemically stable theories. The first is Tait's account of finitism. I argue here that the finitist view does not satisfy the requirement of there being a description of the theory S that the finitist recognizably accepts. Thus, this part of Dean's argument simply does not apply to the version of the ICT that I defend. The second foundational view is Isaacson's thesis that PA is epistemically privileged. I argue that this view does undermine the fully general version of ICT, but that a slightly weaker version is unaffected, and that this weaker version is all an epistemologist would need for the purpose of understanding the epistemology of reflection principles for agents like us.

2. Commitment to Reflection Principles

Dean's statement of ICT quoted above refers to 'various additional statements Δ'. The additional statements I will be concerned with are the *local reflection principles* of a theory S. A local reflection principle for S is any instance of the following schema, for ϕ a sentence in the relevant language:[1]

$$Prov_{\mathsf{S}}(\ulcorner\phi\urcorner) \rightarrow \phi$$

1. This of course assumes that the formal system in question is strong enough to do standard Gödel numbering.

Here $Prov_{\mathsf{S}}$ must be a canonical provability predicate for S. $Con(\mathsf{S})$ is equivalent to the case where $\phi := \bot$. (The notion of a canonical provability predicate is fairly standard, but I will return to it in §2.1 below.)

Why is one justified in believing the local reflection principles? One natural idea is that, in believing S, you are committing yourself to things being as S says they are. But this is exactly what the local reflection principle expresses: if S says that ϕ is so, then ϕ is indeed so. So if you have grounds to believe a theory S, you thereby have grounds to believe its local reflection principles. Indeed, it would be perverse to accept some axioms of S, but not admit that you will accept anything you can prove from those axioms. This point, or at least something like it, is common in the literature.[2] On the other hand, this line of thought, although evocative, is somewhat imprecise. It would be nice to explain more perspicuously exactly *how* and *why* you have grounds to believe the local reflection principle for S whenever you have grounds to believe the axioms of S.

Another way to answer the question of why one is justified in believing the local reflection principles is to appeal to the notion of truth. If you are justified in believing the axioms of S, then you are justified in believing that they are true. Then arguing by induction on the length of proof, you can infer that anything that can be proved is true. So by disquotation, if you can prove ϕ from S, then ϕ. As a way to defend ICT, this strategy has two main drawbacks. First, it involves one in substantive debates about the nature of truth and the proper formal theory of truth. On some popular and attractive theories of truth, the argument will not go through.[3] Second, it is not clear that this strategy actually supports ICT. Since the appeal to a theory of truth makes all instances of the local reflection principle provable via the inductive argument, the commitment to the reflection principle is actually explicit in the acceptance of S plus the truth theory. The relevant epistemological question is then no longer *what is the nature of implicit commitment?* but rather, *what is the epistemic standing of the truth theory?*

I will try to find a way between these two strategies by offering a new argument that is fully precise and explains *why* commitment to a theory also commits you to its reflection principles, but which does not appeal to truth-theoretic notions. The formulation of ICT that I will defend is that if you have mathematical certainty in a theory under a fixed description of its axioms and rules, then you are also in a position to justifiably believe its local reflection principles with mathematical certainty.[4] There are two important aspects to this formulation,

2. See, e.g., Myhill (1960), Feferman (1962), Tennant (2002), Horsten and Leigh (2017).

3. See Field (2006). Dean (2014: 54–61) also contains a long critical discussion of the inductive argument's merits.

4. By *being in a position to justifiably believe* I mean simply that you have evidence available that justifies a belief, so that if you properly formed that belief it would be doxastically justified. To be somewhat more careful, it would be better to state the ICT as saying that if *you are in a position*

namely the notion of mathematical certainty and the fixed description of the theory's axioms and rules. I will comment on each of these before presenting my argument in defense of ICT. I want to flag, however, that while my positive defense of ICT depends on both of these notions, my reply to Dean's objection will not rely on the notion of mathematical certainty. Thus, my reply to Dean's objections in §3 is available even to those who reject the idea of mathematical certainty.

2.1. A Description of Axioms and Rules

I will begin with the less controversial aspect of my formulation of ICT, namely that S must be given under some fixed description of its axioms and rules which the agent in question is able to grasp and assent to. Here is why we need that condition. As a theory, in the technical sense, S is just a set of sentences closed under deduction. But my topic here is epistemological, so I am interested specifically in theories that can be known, or at least believed; and thus we must have some way of grasping or apprehending the set of sentences S. Since we cannot directly apprehend infinite sets of sentences off in Plato's heaven (whatever that would even mean), we must have some more immediate, concrete way of apprehending a theory. Typically, we grasp a set of sentences by means of an axiomatization. But there may be multiple distinct axiomatizations of a given theory, and we may be in a position to justifiably assent to some of those axiomatizations and not others. Thus, we want to hold fixed the particular axiomatization that we are in a position to grasp and accept.

In order to apprehend a finitely axiomatizable theory, it suffices to write down its finitely many axioms. Many theories of interest are not finitely axiomatizable, however. In such cases, we presumably need a *finite description* of the infinite axiomatization, and the standard way of providing such a description is by means of axiom schemas, as with the schema of induction in arithmetic or the schema of separation in set theory. In this case, in order to be said to grasp and accept the axiomatization, the agent should be able to grasp and recognize the schema qua schema; that is, she should recognize its intended range of generality and be disposed to assent to each instance of the schema as such.

This takes us back to the restriction that $Prov_{\mathsf{S}}$ be a canonical provability predicate. Let me detail what I mean by that. I will assume we have some straightforward presentation of the axioms of S as a finite list of axioms and

to have mathematical certainty in a theory under a fixed description of its axioms and rules with mathematical certainty, then you are also in a position to justifiably believe its local reflection principles with mathematical certainty. This gets very wordy, however, and the differences between these two versions will not matter here.

axiom schemas (and rules, if any). For brevity, let S_a denote a particular description of the axioms (and rules, if any) of S. I also assume we have some standard deductive system, and that we have developed a coding apparatus in S capable of representing the usual properties such as *is a formula*, *is a sentence*, etc.

Now a canonical provability predicate is a formula defined in two steps: first, we transcribe directly into the coding apparatus our definition of the axioms of S as being a member of S_a. This is contrasted with 'non-standard' ways of defining the axioms of the theory S, such as by defining S as the maximal consistent subtheory of some set of axioms given by a list. This way, the agent in question is able to grasp and assent to the formalized definition of S_a just as they were able to grasp and assent to the unformalized definition.

The second step is to take this formalized definition of S_a and transcribe directly into the coding apparatus the definition of *proof whose premises are among* S_a in our chosen deductive system. Then a canonical provability predicate is the formula that 'says' *there exists a proof from* S_a *whose conclusion is* ϕ. (Thus, it would be more perspicuous to abbreviate the canonical provability predicate as $Prov_{\mathsf{S}_a}$ rather than $Prov_{\mathsf{S}}$, and I will do so.)

The significance of all this is that, because the agent is able to grasp the theory S under the description S_a, when $Prov_{\mathsf{S}_a}$ is a canonical provability predicate, the agent is also able to recognize that it expresses the property of being provable in S.

2.2. Mathematical Certainty

Let us turn now to mathematical certainty. I will adopt a notion of certainty as the absence of reasons for doubt.[5] On my view, what is distinctive of mathematical certainty, as opposed to practical certainty, empirical certainty, or absolute certainty, concerns the nature of the grounds for doubt. To have mathematical certainty in ϕ is to have no *mathematical* grounds for doubting ϕ. Thus, for instance, I can have mathematical certainty in the claim that $1+1=2$, even if philosophical worries about the existence of abstract objects give me reason to doubt the literal truth of this claim. These grounds of doubt are not mathematical in character. By contrast, an experienced mathematician might have very reliable intuitions about, say, whether a given series converges. On this basis they might justifiably have very high confidence that a series converges. Nevertheless, this intuition, though reliable, is not conclusive, and there may remain some mathematical grounds for doubt that the series converges.

5. This is a standard conception of certainty, being found, for instance, in Giaquinto (2002) among others. An alternative view, found in Frankfurt (1962) and Miller (1978), is to connect certainty in ϕ with a willingness to take risks associated with ϕ. There is a clear affinity between these views, inasmuch as doubts about ϕ will tend to make one hesitate to take on risks associated with ϕ.

There will of course be examples of doubts that are more difficult to classify as mathematical or non-mathematical in character. For instance, consider the debate between intuitionists and classical mathematicians about whether the Law of Excluded Middle holds for arithmetical sentences. Does the viability of intuitionism provide mathematical reason for doubting LEM? Or is it a philosophical reason, being based in the intuitionist's metaphysical views about mathematical objects? In my opinion, these doubts are better classified as philosophical rather than mathematical, but the matter is not entirely clear.[6] As another example, many set-theorists doubt $V = L$ on the grounds that L is much too small and restrictive to comprise the entire universe of set theory. On the one hand, this might seem mathematical, because it is grounded in certain expectations about which mathematical claims should be true.[7] On the other hand, it might seem to be a philosophical doubt because it is based on certain philosophical ideas about how the universe of sets should look.

Other cases are easier to classify, though. For instance, the *P* vs. *NP* problem is a clear example of an open mathematical problem. There are plausibility arguments and some inductive evidence for the claim that $P \neq NP$.[8] But these fall short of conclusive mathematical evidences, and there is clear room for mathematical doubt about *P* vs. *NP*.

In general, wherever we have mathematical arguments that fall short of being dispositive we can find mathematical room for doubt. Plausibility arguments, limiting-case considerations, trained intuition, enumerative induction, and probabilistic methods can all provide mathematical reasons to believe some proposition. But the reasons that these methods provide will not be conclusive and will leave open the possibility of doubt. As a concrete example, a probabilistic primality test can make it very likely that a given number is prime. But the test is merely probabilistic; indeed, it is provable that the test has a certain error rate. Hence, you still have mathematical grounds for doubting that the number is in fact prime.

What matters for present purposes is whether there are some clear, non-trivial examples of theories that we can have mathematical certainty in and which exhibit Gödelian incompleteness. If so, then there are non-trivial examples of theories to which the ICT applies. To be clear, I am not assuming that all of mathematics is certain, and I do not need to take a stand on where or how the

6. Cf. McCarty (2005) who argues that the difference between intuitionists and classicists is mathematical.

7. For instance, if $V = L$, then there is a Σ^1_2 well-ordering of the continuum (Jech 2003: 494). This might seem mathematically doubtful because the continuum is not 'tame' enough to admit such a simple well-ordering.

8. Scott Aaronson summarizes several of these arguments in a blog post, "The Scientific Case for $P \neq NP$," https://www.scottaaronson.com/blog/?p=1720.

line might be drawn between the certain and the less-than-certain (though I will return to this question in §2.4). I am also not committed to any particular view on the nature of mathematical justification. A traditional view held that mathematical axioms must be self-evident; and while this plausibly applies to some axioms such as $0+1=1$, it has more recently become widely acknowledged that holistic, abductive, and pragmatic factors can also play a role in justifying a choice of axioms.[9] Despite invoking a notion of mathematical certainty, I am not reverting to the traditional Euclidean view that all of mathematics consists in deduction from self-evident axioms. I am not even committed to the claim that self-evidence is the source of whatever certainty there is in mathematics. (For all I have said here, it may be possible to obtain mathematical certainty from holistic considerations!) All that I require is the claim that there is some theory which does enjoy mathematical certainty and which exhibits Gödelian incompleteness.

I offer Robinson arithmetic Q as such an example. It is a classic result that Q suffices for proving Gödel's incompleteness theorem.[10] It also seems to me that the axioms of Q are beyond any mathematical doubt. After all, they are only the simplest axioms characterizing 0, 1, $+$, $\times$, and $<$. To the skeptic, I ask: what are your mathematical grounds for doubting the axioms of Q?

2.3. The Argument for ICT

The argument for ICT is very simple. At the heart of the argument is the idea that justification flows through deduction, so that if some evidence provides justification for belief in a set of statements, then that same evidence provides justification for belief in anything those axioms entail. Note that I am *not* assuming that knowledge is closed under deduction, *only* that evidential support flows through deduction. If this were not so, then it would be difficult to explain the role of deduction in our reasoning practices or how it is even *possible* to gain knowledge by inference. This idea is the fundamental motivation for the argument, but the precise claim the argument relies on is this: if ϕ is provable from Δ, and you have some grounds for doubting ϕ, then that same ground provides

9. Shapiro (2009) is a nice account of this shift in view, emphasizing holistic factors in axiom-justification. Maddy (1988a) and Maddy (1988b) are a classic study on abductive and pragmatic factors behind axioms of set theory. Koellner (2006) is a more recent account of work that might be described as providing abductive justification for axioms in set theory. See also Clarke-Doane (2020: ch. 2) for an overview of these issues.

10. See Hájek and Pudlák (1993: III.2.b). While Q provides a nice rhetorical starting point for my argument, I don't see that much would be lost by starting with, say, Elementary Arithmetic (also known as EFA) or some other simple theory.

reason for doubting Δ.[11] It is helpful heuristically to suppose that there exists reason to doubt ϕ just in case there is a plausible scenario where ϕ does not hold. Under this paraphrase, the main premise of the argument can be stated as: if ϕ is provable from Δ, then if there is a plausible scenario where ϕ fails, then there is a plausible scenario where some member of Δ fails.

Now suppose ICT is false, so you have mathematical certainty in a list of axioms S_a, but for some sentence ϕ, you do not have mathematical certainty in the conditional $Prov_{\mathsf{S}_a}(\ulcorner\phi\urcorner) \rightarrow \phi$. Then there are some mathematical reasons for doubting $Prov_{\mathsf{S}_a}(\ulcorner\phi\urcorner) \rightarrow \phi$. This would require a mathematically plausible scenario where ϕ is indeed provable from the axioms S_a, but nevertheless ϕ fails to hold. But this, I claim, is ruled out by the fact that you have mathematical certainty in S_a. I take it as obvious that there is no mathematically plausible scenario in which the axioms S_a all hold but some provable consequence ϕ does not hold. And since S_a are all mathematically certain, there is no mathematically plausible scenario where S_a do not all hold. And thus there is no mathematically plausible scenario where ϕ is a provable consequence of S_a but does not hold.

There is a clear connection between this argument and the truth-theoretic argument for ICT sketched earlier, and one might object that my argument actually presupposes the truth-theoretic argument. While I have avoided use of the word 'true', I referred to things that hold or do not hold. Saying ϕ holds is plausibly equivalent to saying that ϕ is true, and I assumed implicitly that when some sentences Δ hold, so do any of their consequences. In other words, the objection goes, my argument requires all of the assumptions of the truth-theoretic argument and hence does not provide any new defense of ICT.

In response to this objection, I would point out that the talk of sentences holding in some scenarios is a vivid way of grasping the argument, but is ultimately dispensible. The essential argument can be expressed entirely in terms of doubt and certainty. (This point is analogous to the common observation that it is often helpful to reason about modal matters using possible worlds, but this can ultimately be dispensed with in favor of a primitive modal operator.) Write $\mathcal{C}\psi$ to mean that ψ is mathematically certain. Then the two key premises of the argument are:

11. One might object here that sometimes the doubt in ϕ would justify doubting classical logic rather than the premises Δ. I would reply, however, that one need not *believe* that Δ entails ϕ in order for doubts about ϕ to provide grounds for doubting Δ as well. In general, one need not *believe* that one's evidence justifies adopting some doxastic attitude in order for one's evidence to *actually* justify adopting that attitude. So even if you have doubts about classical logic, that does not mean that the strictures of classical logic do not apply to you. More generally, standards of rationality are binding even on agents that do not believe in those standards. But a full defense of this view gets entangled with debates about externalism, akrasia, and higher-order evidence, and is thus outside the scope of this paper. Authors that defend similar views include Titelbaum (2015) and Littlejohn (2018); see also Weatherson (2019) for another, much stronger, view in this vein. (Thanks here to an anonymous referee.)

- $\mathcal{C}\neg(\mathsf{S}_\mathsf{a} \wedge Prov_{\mathsf{S}_\mathsf{a}}(\ulcorner\phi\urcorner) \wedge \neg\phi)$.
- $\mathcal{C}(\psi \to \theta) \to (\mathcal{C}\psi \to \mathcal{C}\theta)$.

From these two premises it is easy to derive ICT.[12]

Thus, my argument does not clearly rely on the same assumptions as the truth-theoretic argument. It is true, however, that there is some similarity between the two arguments, in that both arguments rely on some property being preserved under deduction. For the truth-theoretic argument, this property is truth, while in my argument the property is mathematical certainty. At the same time, there is an important difference between the arguments in their dialectical role and in who is supposed to grasp them. Recall the structure of the truth-theoretic argument: if the axioms of some theory are true, then by induction on the length of proof we can argue that anything provable from them must also be true. So we can conclude that the local reflection principles are true. On its face this says nothing about the epistemic standing of the reflection principles. To infer that some agent is committed to the local reflection principles we need further assumptions. The most natural way to fill in this gap is with the two assumptions that, first, the agent is explicitly or implicitly justified in believing the relevant truth theory, and, second, one is justified in believing the consequences of whatever one is justified in believing.[13]

The important difference between this line of reasoning and my own argument is that this line of reasoning requires the agent in question to have some justification for believing the truth theory. Furthermore, in order for the agent to believe the reflection principles with doxastic justification, they presumably have to be able to *perform* the relevant deduction of the reflection principle from the truth theory. If that is correct, then in order for the agent to be in a position to justifiably believe the reflection principles, they have to be able to *grasp* and *use* the truth theory. By contrast, my argument solely concerns the *structure* of epistemic reasons—reasons for believing and doubting propositions. The agent in question does not have to be able to articulate those reasons or the structure between them, provided only that they are appropriately responsive to those reasons in their epistemic behavior. This is why the commitment in question is *implicit*: the agent implicitly has the justification to believe the reflection principles just in virtue of the nature of their epistemic reasons.

I have defended the following version of the implicit commitment thesis: Whenever one is mathematically certain of some theory, under a fixed description

12. From the first premise $\mathcal{C}\neg(\mathsf{S}_\mathsf{a} \wedge Prov_{\mathsf{S}_\mathsf{a}}(\ulcorner\phi\urcorner) \wedge \neg\phi)$ we get $\mathcal{C}(\mathsf{S}_\mathsf{a} \to (Prov_{\mathsf{S}_\mathsf{a}}(\ulcorner\phi\urcorner) \to \phi))$. (This follows from the second premise on the assumption that tautologies are mathematically certain, but that is hardly objectionable.) We then get $\mathcal{C}\mathsf{S}_\mathsf{a} \to \mathcal{C}(Prov_{\mathsf{S}_\mathsf{a}}(\ulcorner\phi\urcorner) \to \phi)$ by the second premise, and this is the ICT.

13. In a mathematical context we may plausibly assume there are no relevant defeaters or other complicating factors that would undermine this closure principle.

S_a of its axioms and rules, one is also in a position to be mathematically certain of the local reflection principles $Prov_{\mathsf{S}_a}(\ulcorner\phi\urcorner) \rightarrow \phi$. In the following subsections, I will address a few objections and discuss possible extensions of my argument.

2.4. Is the ICT Trivial?

One possible response to this argument is to grant its cogency, but object that it is not interesting because it only applies to comparatively weak theories for which we already have independent means of establishing the reflection principles. For instance, say that we can be mathematically certain of PA. Then my argument would show that we can be mathematically certain of the theory $\mathsf{PA} + \{Prov_{\mathsf{PA}}(\ulcorner\phi\urcorner) \rightarrow \phi : \phi \in \mathcal{L}_{\mathsf{PA}}\}$. But this theory is not that much stronger than PA itself, and is properly subsumed by other theories, such as ZFC, that we already have good reason to believe. So, one may object, we have not made any epistemic progress.

Let us grant for the sake of argument that whenever my version of the ICT guarantees us justification for some reflection principles, we also have independent reason to believe those reflection principles. Even so, we can reply to this objection by noting that it is generally of epistemological interest to note what grounds there are for a certain belief even when some of those grounds might be redundant. For instance, my visual perception of an event does not make my auditory perception of the same event epistemologically irrelevant or uninteresting.

Building on this observation, we can note two things that are distinctive of the justification guaranteed by the ICT. The first is the strength of the justification guaranteed by the ICT. In the version I have defended, the ICT says not merely that we have some justification for believing reflection principles, but that we can believe them with *certainty*. While we might be able to prove the reflection principles for a theory S in some stronger theory S', there is no guarantee that S' is mathematically certain. For instance, one might think that Zermelo set theory enjoys mathematical certainty, but that the Axiom of Replacement is not certain. One can prove the consistency of Zermelo set theory in ZFC, but if the relevant instances of Replacement are not mathematically certain, then the proof will not warrant mathematical certainty in its conclusion.

The second distinctive feature of the ICT is the *structure* of the justification it guarantees. The ICT, as I have defended it, shows that mathematical certainty has a self-reproducing character. When a theory S enjoys mathematical certainty, then the ICT tells us that so does a stronger theory S' obtained by adding to S its local reflection principles. Then S' can be extended to another mathematically certain theory by adding *its* reflection principles, and so on.

This second observation has an interesting corollary, namely that, under some general assumptions, we cannot characterize with mathematical certainty the portion of mathematics that enjoys mathematical certainty. Specifically, it is impossible to be mathematically certain that a recursive axiom system includes all and only the mathematically certain propositions. For suppose that we had some recursive axiomatization A, and we were mathematically certain that A entailed all and only the sentences that enjoyed mathematical certainty. Then the ICT would say that we can also be certain of $Con(\mathsf{A})$, along with the other local reflection principles. But since A is supposed to be recursive, and given the definition of a canonical provability predicate above, $Prov_{\mathsf{A}}(x)$ will be a Σ_1 formula and hence the incompleteness theorem entails that $\mathsf{A} \nvdash Con(\mathsf{A})$. So, although we can be certain of $Con(\mathsf{A})$, it is not entailed by A itself, contradicting the assumption that A entails all mathematically certain sentences.

There is no guarantee that the *only* way for us to apprehend theories is by recursive axiomatizations. So this observation does not guarantee that there is *no* way to characterize the mathematically certain propositions with mathematical certainty. But any theory that we apprehend by a finite list of axioms and schemas will fit this description.[14] Since this includes all familiar mathematical theories, we can reasonably summarize the conclusion here by saying that if any theory T captured, with mathematical certainty, exactly the mathematically certain sentences, then T must be in some sense *unfamiliar*.

2.5. Commitments of Uncertain Theories

Although my version of the ICT is not trivial, one might still think that it does not go far enough. For simplicity, I will mostly focus here on the single instance of local reflection $Con(\mathsf{S}_a)$, though my comments will generalize. One might think that accepting *any* theory commits one to the consistency of that theory, regardless of whether that theory is certain or not. So even if your belief in S_a falls short of mathematical certainty, you still have a commitment to $Con(\mathsf{S}_a)$. After all, if S_a were inconsistent, then anything at all could be derived from S_a, which would undermine the acceptance of S_a as a mathematical theory.

There are two ways of fleshing out this essential idea. The first way is to observe that accepting a theory has a pragmatic aspect as well as a doxastic aspect. When you accept a theory S, you are not only assenting to its axioms S_a, you are adopting a commitment to rely on S in your reasoning. (You are possibly also adopting other commitments such as regarding open questions about S as being legitimate issues of mathematical inquiry, though this is not important for

14. Assuming that it is decidable which instances of the schema are to be counted as axioms.

what follows.)[15] In order for this practical commitment to be reasonable, though, reliance on S should not render all reasoning trivial. And if S were inconsistent, all reasoning would be trivial because everything would follow from S. So accepting S rationally presupposes the consistency of S.

This line of reasoning is fine as far as it goes. It has two shortcomings, however. The first shortcoming is that, even by the lights of this pragmatic argument, it is not clear that if S were inconsistent then all reasoning would be trivial. For instance, 'almost-consistent' theories are theories that may be inconsistent, but where the only proofs of inconsistency are unfeasibly long.[16] In an almost consistent theory, any reasoning that one could practically undertake would be nontrivial and, depending on the example, may even be perfectly reliable.[17] This is a comparatively minor point, however. The more significant shortcoming with this line of reasoning is that it is entirely *pragmatic*. It shows that as a matter of practical reason, the acceptance of a theory presupposes its consistency. In this paper, however, I am concerned with the *epistemic* standing of reflection principles. Showing that our practical commitments presuppose $\mathit{Con}(\mathsf{S}_a)$ does not show that we have any epistemic reason to believe $\mathit{Con}(\mathsf{S}_a)$.

As I said, there are two ways of fleshing out the original objection we started with. If the first way fails to bridge the gap between the practical standing and the epistemic standing of $\mathit{Con}(\mathsf{S}_a)$, the second way tries to overcome this gap by taking some ideas of the first approach but looking at them in an epistemic light. Similar to the first approach, we begin by noting that believing a theory S is not an isolated cognitive act or state, but it is part of a larger cognitive project of mathematical inquiry. And it is not just a pragmatic requirement, but indeed an (epistemically) rationally required presupposition of that cognitive practice that our beliefs be consistent. (Of course, we do not always live up to this requirement, but that only shows that we are not perfectly rational, not that the requirement is not genuine.) And when ϕ is some rational presupposition of a cognitive project, the argument goes, an agent engaged in that project has a default rational *entitlement* to believe ϕ.[18] Thus, being engaged in a project of mathematical inquiry, if one believes S, then one is rationally entitled to believe $\mathit{Con}(\mathsf{S}_a)$.[19]

15. It is generally thought that acceptance does not require full-blooded belief; my use of 'assent' in this context should accordingly be read in a non-beliefy way. Cf. Cohen (1989) on the notion of acceptance; van Fraassen (1980) is a classic study of the role of acceptance in scientific inquiry.

16. This idea comes from Parikh (1971).

17. To be more exact, we should really speak of reasoning with a particular presentation of a theory. After all, we reason with some representation of a theory, not with the theory qua infinite set of sentences. This is just terminology, however.

18. This idea of entitlement comes from Burge (2003) and Wright (2004).

19. Horsten and Leigh (2017) and Horsten (2021) endorse the view that one has an entitlement to accept reflection principles of theories one accepts. I do not claim to have captured all the subtleties of their views here.

This argument is again fine as far as it goes. Entitlement, however, is generally taken to be a weaker notion than justification; so while this argument applies to a wider range of theories than my argument does, it yields a strictly weaker conclusion about them. The notion of entitlement is also more controversial than that of justification, so there is interest in determining the extent to which one has full-fledged justification to believe reflection principles rather than mere default entitlement.

This raises the question of whether my argument that we have *justification* to believe reflection principles can be extended beyond those theories that are mathematically certain. I cannot rule out that there is *some* argument for this stronger claim,[20] but I am doubtful that it would proceed similarly to the argument I have given here. Here is why.[21] The motivation for the argument I gave above was that justification flows through deduction. Without such an assumption, it is not clear my argument even gets off the ground. But that assumption also motivates the following claim, where $J\psi$ is to be read as "you are in a position to justifiably believe ψ":

$$J(\mathsf{S}_a) \rightarrow (Prov_{\mathsf{S}_a}(\ulcorner \phi \urcorner) \rightarrow J\phi)$$

After all, if you are in a position to justifiably believe S_a, then by competently performing the deduction of ϕ from S_a you would put yourself in a position to justifiably believe ϕ. Taking the instance of $\phi := 0 = 1$ and using propositional logic, this gives:

$$\neg J(0 = 1) \rightarrow (J(\mathsf{S}_a) \rightarrow \neg Prov_{\mathsf{S}_a}(\ulcorner 0 = 1 \urcorner))$$

Assuming, quite reasonably, that one is never in a position to justifiably believe $0 = 1$, by modus ponens we have:

$$J(\mathsf{S}_a) \rightarrow \neg Prov_{\mathsf{S}_a}(\ulcorner 0 = 1 \urcorner)$$

That is, the relevant notion of justification requires that it be *impossible* to justifiably believe an inconsistent theory. This requires a very strong notion of justification. So even if this includes justification that is somewhat short of mathematical certainty, it is unlikely that the argument I have given, or one akin to it, would apply to theories that enjoy a significantly weaker form of justification.

20. Indeed, I find this stronger claim rather attractive, so I hope there is such an argument.
21. I owe the following argument to an anonymous referee.

2.6. The Need to Learn Gödel Coding?

I have so far argued at an informal level, ignoring the fact the local reflection principles are formal sentences in the language of S. Might that make a difference to their epistemic standing? For my part, I cannot see why it should make any more difference to the epistemic standing of a belief that it be expressed in the language of arithmetic (say) than if it were expressed in the language of French (say). If one does not know French, then one may need to acquire new conceptual resources to appreciate that they have reason to believe *le chat est noir*. But if one already knows French, then seeing the black cat will suffice to give them reason to believe *le chat est noir*. Moreover, in learning French, one has not fundamentally altered their epistemic state, at least as far as the black cat is concerned—one simply put oneself in a position to take advantage of their epistemic reasons in favor of assenting to 'le chat est noir'. Similarly, if one were not familiar with how to use the coding apparatus to represent the syntax of S within S, then they may require some new conceptual resources to appreciate that they have reason to believe $Prov_{S_a}(\ulcorner\phi\urcorner) \rightarrow \phi$.[22] But since we are assuming that $Prov_{S_a}$ is a canonical provability predicate, once one learned how to transcribe an ordinary finite list of axioms or axiom schemas and the definition of a proof into the coding apparatus, one would then be in a position to appreciate that the formal sentence $Prov_{S_a}(\ulcorner\phi\urcorner) \rightarrow \phi$ expresses the same thing as *if there is an* S*-proof of* ϕ*, then* ϕ. Of course, learning how to use a coding apparatus can be a non-trivial task. But then so can learning French. And just as learning French does not affect the epistemic standing of one's belief that the cat is black, but only gives one a new way of expressing that belief, so learning the formal coding apparatus for S does not affect the epistemic standing of one's belief that if there is an S-proof of ϕ then ϕ, but merely gives one a new way of expressing that belief.

A related objection one often hears is that there is an important difference between an informal claim about the consistency of a theory S and the formal consistency sentence *Con*(S), because the formal arithmetized sentence is *about* numbers rather than being *about* the consistency a formal theory. This ostensible difference in subject matter is then alleged to be of some epistemic import. If so, however, the point would seem to generalize beyond the formalization of syntax to other uses of formalization. For instance, graph theory can be formalized in arithmetic. Is there some important difference between the *epistemic* standing of sentences in graph-theoretic language and sentences in the language of arithmetic that formalize those same graph-theoretic claims? I can see no basis for thinking so.

22. See also Horsten (2021: 10–11) in this connection. The fact that $Prov_S$ is a canonical provability predicate is important for both Horsten and me.

2.7. Other Reflection Principles?

I have cast my argument in terms of local reflection principles. Does it extend also to the uniform reflection principle:

$$\forall x Prov_{\mathsf{S}_a}(\ulcorner \phi(\dot{x}) \urcorner) \rightarrow \forall x \phi(x) \ ?$$

Or to the global reflection principle, where $Sent(x)$ means that x is a (code of a) sentence and Tr is a truth predicate:

$$\forall x[Sent(x) \wedge Prov_{\mathsf{S}_a}(x) \rightarrow Tr(x)] \ ?$$

Nothing I have said requires me to either accept or deny ICT applied to uniform or global reflection principles, and I am happy to remain officially neutral. I will, however, note a few complications that would arise in extending the above argument to apply to global and uniform reflection.

Obviously, the global reflection principle only carries any substance when there is a truth theory in the background.[23] Thus, extending the implicit commitment argument to apply to the global reflection principle involves taking a stand in the fraught debate surrounding theories of truth.

A second point about the global reflection principle is that it seems to require more conceptual resources on the part of the believing agent. For an agent to be in a position to know the global reflection principle for S, they must have the concept of truth. This may or may not be a substantial assumption, depending on the nature of truth. But I have tried to show how it can be avoided by explaining implicit commitment to local reflection principles without appealing to the concept of truth.

On the question of uniform reflection principles: their formulation assumes that every object in the domain has a name (this is a background assumption of the dot-notation). This is of course problematic for most theories other than arithmetic. So while my argument for implicit commitment to local reflection principles is fully general, the potential scope of a similar argument for implicit commitment to uniform reflection principles would be severely restricted. Even in the restricted case of arithmetic, adapting the argument above to show that belief in, say, PA incurs a commitment to the uniform reflection principles would seem to require the assumption that the believing agent knows that every object in the domain has a name.[24] If the agent's names for numbers are just the usual

23. Otherwise, it would be consistent to assume that all and only the provable sentences were true, and hence the result of adding the global reflection principle to S would be conservative over S.

24. Or, to hedge a bit, perhaps it would require that the agent be in a position to know that every object has a name.

numerals, this commitment would apparently allow the agent to rule out non-standard models. As with a theory of truth, this may or may not be a substantial assumption. There may or may not be a plausible account of how the agent is able to distinguish standard and non-standard models of arithmetic. I am simply observing that extra complications arise in trying to extend my argument to global or uniform reflection principles.[25]

3. Dean's Objection to Implicit Commitment

The claim that accepting a theory S incurs an implicit commitment to reflection principles for S is a fairly common one, but recently Dean (2014) has made an interesting case against this claim. Dean's main argument is that there are foundational views of number theory that are epistemically stable in the following sense: from the perspective of such a foundational view, that view itself is warranted but no stronger view is warranted.[26] Suppose that an epistemically stable foundational view F is adequately captured by a system of arithmetic S; then the reflection principles for S are not warranted according to F, since they outstrip the system S which, by hypothesis, adequately captures the foundational view F. Thus, acceptance of such theories of arithmetic does not necessarily incur an implicit commitment to their reflection principles. Dean provides two case studies of such foundational views: finitism and Isaacson's thesis.

An interesting feature of Dean's argument is that it does not assume anything about the nature of epistemic warrant. His objections, if successful, would undermine both the sort of view I have defended, on which we have genuine justification to believe reflection principles, as well as weaker views according to which we are merely entitled to believe reflection principles as sketched in Section 2.5, as well as any possible views in between these. In replying to Dean, therefore, I will try to be neutral about the nature and source of our warrant for believing reflection principles. They may be justified with mathematical certainty, they may enjoy some slightly weaker justification, they may only enjoy

25. Following a very different strategy than the one I have developed here, Fischer (2021) develops an intriguing argument that the uniform reflection principle can be justified by a conception of the natural numbers as an inductive structure.

26. Dean's paper also includes two other considerations against ICT. The first is that reflection principles are generally equivalent to strong forms of transfinite induction. Thus, an implicit commitment to reflection is tantamount to an implicit commitment to a strong principle of transfinite induction, which might strike one as implausible. I do not find this consequence terribly implausible myself, however. And since this disagreement comes down to a conflict of intuitions, it is unlikely to be a productive line of debate. Second, Dean notes various difficulties that arise if one tries to justify ICT by appeal to a truth theory. Since I have not appealed to a truth theory, I do not have to face these difficulties.

entitlement. The goal is that my reply to Dean be available not only to myself, but also to others who adopt a different version of the ICT. Thus I will not rely on the assumption that the ICT applies only to theories believed with mathematical certainty. I will, however, rely on the assumption that the ICT only applies when an agent grasps a theory under a particular description. Indeed, this will be the heart of my response to Dean's case study of finitism. Since this is a very general feature of how we grasp we mathematical theories, this assumption should not diminish the generality of my reply to Dean.

I will begin with the example of finitism. While this example does reveal problems for the version of the ICT as Dean states it (quoted above in the introduction), I will argue that these problems do not affect the the version of the ICT that I have stated and defended. Then I will consider the example of Isaacson's thesis; I grant that this case study undermines the most general version of the implicit commitment thesis but argue that a slightly weaker version survives unscathed.[27]

3.1. Finitism

With finitism, Dean has in mind the influential analysis of Tait (1981). Tait aims to delimit the functions and number-theoretic statements which are meaningful from the finitist point of view. He argues that the general conception of a function as an arbitrary mapping between objects of a domain is not finitistically meaningful. Rather, we have to ask about the specific mappings that can be finitistically constructed on a specific finitistically acceptable domain. Thus Tait aims to give an account of the functions that can be finitistically constructed on the basis of the finitist view of the natural numbers.

Tait's account begins from the idea that, according to the finitist, numbers are apprehended as (finite) sequences; the concept Number, then, is the "generic form of a finite sequence" (Tait 1981: 530). Accordingly, the basic operation implicit in the finitist picture of the natural numbers is *iteration*, as we obtain one finite sequence from another by the mapping $n \mapsto n+1$. On this basis, Tait argues that the finitistically acceptable functions are exactly the primitive recursive

27. Thus my response to Dean's examples is markedly different from Nicolai and Piazza (2019). Nicolai and Piazza grant that Dean's case studies successfully undermine the claim that implicit commitment includes commitment to further object-language sentences that are independent of the theory in question. They go on to distinguish another, meta-theoretic sense of implicit commitment, however, which they dub the *semantic core* of implicit commitment. Roughly, the semantic core is a compositional theory of truth which Nicolai and Piazza argue one is implicitly committed to. Since the truth theory is conservative over any base theory extending Elementary Arithmetic, this is compatible with the claim that implicit commitment does not always include commitment to further sentences in the language of the base theory.

functions. From this claim, Tait then further argues that the finitistically acceptable theorems of arithmetic are precisely those of primitive recursive arithmetic (PRA).

However, the coincidence of primitive recursive arithmetic and finitist arithmetic is not something the finitist can recognize. Having no general conception of a function, the finitist cannot understand primitive recursion as a higher-level operation of functions; they can merely apply the schema of recursion to particular functions that are already given to them. As Tait explains, "For the finitist to recognize the validity of primitive recursive arithmetic, he must recognize the general validity of definition of functions by primitive recursion. But he cannot even formulate this since it involves the notion of function" (1981: 545). Since the finitist is not able to recognize the validity of PRA, they are also not in a position to justifiably accept the reflection principle $Prov_{\mathsf{PRA}}(\ulcorner\phi\urcorner) \rightarrow \phi$.[28]

The point here is that, although the finitist can justifiably accept each axiom of PRA, there is no means by which the finitist can apprehend the whole of PRA. Accordingly, there is no fixed description of the axioms and rules of PRA such that the finitist is in a position to justifiably accept the whole of PRA under that description. My claim about the implicit commitment to reflection principles, however, applied only when one is in a position to justifiably accept a mathematical theory under a fixed description of its axioms and rules. Thus, Dean's argument from the epistemic stability of finitism does not undermine the implicit commitment claim as I have formulated it here. Moreover, my formulation of the ICT is independently motivated by the considerations about how we apprehend and accept theories. If I had simply added an extra condition to the ICT to avoid Dean's counterexample, that would be ad hoc. But I have shown that there is an attractive, well-motivated version of the ICT to which this first counterexample of Dean's does not apply.

3.2. Isaacson's Thesis

Let us turn now to Dean's second case study, Isaacson's thesis. Isaacson argues for the view that PA "occupies an intrinsic, conceptually well-defined region of arithmetical truth" (1987: 147), or as he phrases his view later in that paper,

28. On Tait's account (following Hilbert 1925), the finitist also does not recognize quantifiers. Thus, one might worry that even a reflection principle for a weaker theory such as Robinson's Q would be incomprehensible for the finitist; after all, the provability predicate is a Σ_1 formula. Whether the finitist's rejection of quantifiers is well-motivated is controversial (Incurvati 2015). At any rate the reflection principle $Prov_{\mathsf{Q}}(\ulcorner\phi\urcorner) \rightarrow \phi$ is Π_1, and hence equivalent to a quantifier-free formula, and even Hilbert (1925) is comfortable with a generality interpretation of free variables. But these issues are orthogonal to the main thrust of my reply to Dean.

"[Peano Arithmetic] is complete with respect to purely arithmetical truth" (1987: 166). In a subsequent paper, the epistemological character of Isaacson's thesis comes out more clearly, where he says that for a statement to belong to this distinguished region of arithmetical truths and falsehoods—in short, to be 'arithmetical'—that statement must be expressed in the language of arithmetic and further "that its truth or falsity be perceivable directly on the basis of an articulation of our grasp of the fundamental nature and structure of the natural numbers, or directly from statements which themselves are arithmetical" (Isaacson 1992: 95). This is clearly an epistemological criterion for being arithmetical, in Isaacson's sense of the word. Furthermore, on this characterization of arithmeticity, Isaacson's thesis would entail that PA is epistemically stable in Dean's sense, because our grasp of the fundamental structure of the natural numbers would justify accepting PA but nothing further.

Isaacson offers several considerations in favor of the view that PA occupies a distinguished region of arithmetical truth, drawing on results such as the categoricity of second-order PA and the bi-interpretability of PA with the theory of hereditarily finite sets. Of more direct relevance to present concern, however, is why Isaacson thinks that Gödel sentences and reflection principles are *excluded* from this conceptually distinguished region of arithmetical truth.

Isaacson does not mention local reflection principles in general, but he does discuss the case of the Gödel sentence that expresses its own unprovability. So why does Isaacson think that our grasp of the natural number structure does not by itself justify accepting a Gödel sentence? In considering the possibility of adding a Gödel sentence G as an axiom to PA, Isaacson (1987: 159) writes:

> Such a move would be unnatural. An axiom in this context should be an evident truth, in the terms in which it is expressed. But the truth of this statement, as a statement of arithmetic, is not directly perceivable. $\mathsf{PA} + G$ would not constitute, in this way, a purely arithmetical extension of PA. The Gödel sentence thus offers an instance of the general thesis of this paper that any axiomatic extension of Peano Arithmetic must be motivated by considerations for establishing its truth which rely essentially on non-arithmetical notions.

What are the 'non-arithmetical notions' that are involved in establishing the truth of the Gödel sentence? The answer is clearer in the later paper, where he writes that accepting such sentences would require hidden 'higher-order' concepts:

> In the case of the Gödel sentence for Peano arithmetic, the hidden concepts are [1] provability in the formal system of Peano arithmetic and, most crucially, [2] consistency of Peano arithmetic. That is, to perceive the truth of

> the Gödel sentence (presented purely in the first-order language of arithmetic) we must [1] understand that it expresses the condition that this sentence is not provable in this given formal system and [2] see that this formal system is consistent. (Isaacson 1992: 96; numbering added for clarity)

Isaacson's point here is that we do not perceive the truth of the Gödel sentence G "directly on the basis of an articulation of our grasp of the fundamental nature and structure of the natural numbers, or directly from statements which themselves are arithmetical", as we would have to for G to count as arithmetical in Isaacson's sense. Rather, our belief in G relies on our grasp of the concept of provability in PA and some beliefs about what is provable. We need something beyond a mere grasp of the natural numbers to perceive the truth of G.

Following Isaacson's analysis of the basis for belief in G, we can say something similar about what is involved in accepting a local reflection principle: to perceive the truth of the reflection principle $Prov_{\mathsf{PA}}(\ulcorner\phi\urcorner) \rightarrow \phi$ we need three things. We need to have the abstract concept of proof, we need to be able to grasp the axioms of PA as a collection of statements we accept, and we need to be able to recognize that $Prov_{\mathsf{PA}}$ expresses the abstract concept of proof from the axioms of PA.

Now, I was at pains to emphasize above that for you to be justified in accepting a reflection principle $Prov_{\mathsf{S}}(\ulcorner\phi\urcorner) \rightarrow \phi$, the description of the axioms of S—what I referred to as S_a—must be one that you can recognize as a body of statements you accept. I also argued that if $Prov_{\mathsf{S}_a}$ is a canonical provability predicate, then one *will* be able to recognize that $Prov_{\mathsf{S}_a}$ expresses the abstract concept of proof from the axioms of S. On these points, then, I agree with the Isaacsonian analysis, but this does not undermine ICT as I have defended it.

However, with the claim that we also need the abstract concept of proof, I must admit that the Isaacsonian has a point. This is a relatively sophisticated, high-level concept, and it is possible that one might have a sufficiently developed conception of the natural numbers to justify accepting the axioms of PA, but not yet have the general, abstract concept of proof. For instance, it is conceivable that a very cognitively advanced animal would grasp the axioms of PA and be able to perform reasoning with them without having the concept of a proof in general. To this extent, then, Isaacson is right that the local reflection principles for PA are not justified directly and exclusively on the basis of our grasp of the natural number structure, and hence that local reflection principles are not purely arithmetical in his sense.

Thus I grant that Isaacson's thesis does undermine the following interpretation of ICT:

> Any creature whatsoever who warrantedly accepts a theory S under a recognizable description of its axioms is thereby in a position to warrantedly accept the local reflection principles for S.

On the other hand, Isaacson's thesis does not undermine this interpretation of ICT:

> Any creature who has the general concept of proof and who warrantedly accepts a theory S under a recognizable description of its axioms is thereby in a position to warrantedly accept the local reflection principles for S.

This is a slightly weaker, but still quite interesting thesis. And humans generally do have the abstract concept of proof. Or, to hedge somewhat, humans are generally capable of acquiring the abstract concept of proof. Acquiring this concept can be a difficult task, as witnessed by the training it takes to impart the concept of a proof to undergraduate math students. On the other hand, the concept is not outside the ken of human cognitive capacities, as witnessed by the fact that it is acquired by thousands of undergraduate math students every year. So we can paraphrase this final version of ICT as saying that any creature like us who recognizably accepts a theory S is thereby implicitly committed to further sentences not deductively entailed by S.

Since this weaker thesis has a slightly more restricted range of applicability, it tells us less about the epistemology of mathematics in full generality. But this, I want to suggest, is not much loss. There is certainly interest in studying abstract epistemological questions in full generality. For instance, one might study rationality as such, or one might try to give a fully general conceptual analysis of knowledge. And likewise there is interest in asking about the epistemology of mathematics in full generality, as it applies to all creatures capable of mathematical knowledge. But most epistemological questions are more parochial. For instance, the epistemology of perception is going to depend on what a given creature's perceptual apparatus is like. What counts as a reasonable failure of logical omniscience for non-ideal agents is going to depend on what the particular non-ideal agents in question are like. Similarly, if we are interested in better understanding *our* mathematical knowledge, we will focus on the epistemology of mathematics for agents like us. And for that purpose, ICT as restricted to agents like us is all that we really need.

On reflection, it should not be surprising that Isaacson's thesis says little about the epistemology of actual agents like us. After all, Isaacson's thesis concerns only what is implicit in a grasp of the fundamental nature and structure of the natural numbers. But any actual agent will bring numerous other concepts and capacities to bear on their mathematical knowledge and there is no reason to expect that those other concepts and capacities cannot be combined with a grasp of the natural numbers in a way that yields new mathematical fruit.

4. Conclusion

I have offered a defense of ICT including two components. First, I developed a positive argument for ICT which avoided both the essential imprecision of many previous defenses and the appeal to a truth theory as in other previous defenses. This argument for ICT applies to theories that we believe with mathematical certainty, under a fixed description of their axioms and rules.

The second component in my defense of ICT is the response to Dean's objection to ICT. Unlike my positive defense of ICT, this portion of my argument does not rely on the notion of mathematical certainty. Relying only on the assumption that a mathematical theory is apprehended by means of a fixed description of its axioms, the reply to Dean is available to a wide range of supporters of various implicit commitment theses. I showed that his first case study, Tait's thesis on finitism, does not meet the condition of there being a fixed description of the axioms that a finitist can recognizably accept. And while his second case study, Isaacson's Thesis, does undermine a fully general ICT, a weaker version of ICT restricted to agents like us is unaffected by that argument. This restricted version still holds significant epistemic interest.

Acknowledgments

Precursors of this paper go back several years, and I am grateful to the various people who have discussed these issues with me, including Steve Dalglish, Chris Pincock, Stewart Shapiro, Neil Tennant, and Dan Waxman. Thanks are also due to several referees for both *Ergo* and other journals, whose comments led to an improved paper.

References

Burge, Tyler (2003). Perceptual Entitlement. *Philosophy and Phenomenological Research*, *67*(3), 503–48.

Clarke-Doane, Justin (2020). *Mathematics and Morality*. Oxford University Press.

Cohen, L. Jonathan (1989). Belief and Acceptance. *Mind*, *98*(391), 367–89.

Dean, Walter (2014). Arithmetical Reflection and the Provability of Soundness. *Philosophia Mathematica*, *23*(1), 31–64.

Feferman, Solomon (1962). Transfinite Recursive Progressions of Axiomatic Theories. *Journal of Symbolic Logic*, *27*(3), 259–316.

Field, Hartry (2006). Truth and the Unprovability of Consistency. *Mind*, *115*(459), 567–605.

Fischer, Martin (2021). Another Look at Reflection. *Erkenntnis*. Advance online publication. https://doi.org/10.1007/s10670-020-00363-9

Frankfurt, Harry (1962). Philosophical Certainty. *Philosophical Review*, *71*(3), 303–27.
Giaquinto, Marcus (2002). *The Search for Certainty*. Oxford University Press.
Hájek, Petr and Pavel Pudlák (1993). *Metamathematics of First-Order Arithmetic*. Springer.
Halbach, Volker (2011). *Axiomatic Theories of Truth*. Cambridge University Press.
Hilbert, David (1925). On the Infinite. In Jan van Heijenoort (Ed.), *From Frege to Gödel: A Sourcebook in Mathematical Logic* (367–92). Harvard University Press.
Horsten, Leon (2021). On Reflection. *Philosophical Quarterly*, *71*(4), 1–20.
Horsten, Leon and Graham Leigh (2017). Truth is Simple. *Mind*, *126*(501), 195–232.
Incurvati, Luca (2015). On the Concept of Finitism. *Synthese*, *192*(8), 2413–36.
Isaacson, Daniel (1987). Arithmetical Truth and Hidden Higher-Order Concepts. In The Paris Logic Group (Ed.), *Logic Colloquium '85* (147–69), Vol. 122 of *Studies in Logic and Foundations of Mathematics*. Elsevier.
Isaacson, Daniel (1992). Some Considerations on Arithmetical Truth and the ω-rule. In Michael Detlefsen (Ed.), *Proof, Logic, and Formalization* (94–138). Routledge.
Jech, Thomas (2003). *Set Theory* (3rd ed.). Springer.
Koellner, Peter (2006). On the Question of Absolute Undecidability. *Philosophia Mathematica*, *14*(2), 153–88.
Littlejohn, Clayton (2018). Stop Making Sense? On a Puzzle about Rationality. *Philosophy and Phenomenological Research*, *96*(2), 257–72.
Maddy, Penelope (1988a). Believing the Axioms I. *Journal of Symbolic Logic*, *53*(2), 481–511.
Maddy, Penelope (1988b). Believing the Axioms II. *Journal of Symbolic Logic*, *53*(3), 736–64.
McCarty, D. C. (2005). Intuitionism in Mathematics. In Stewart Shapiro (Ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic* (356–86). Oxford University Press.
Miller, Richard W. (1978). Absolute Certainty. *Mind*, *87*(345), 46–65.
Myhill, John (1960). Some Remarks on the Notion of Proof. *Journal of Philosophy*, *57*(14), 461–71.
Nicolai, Carlo and Mario Piazza (2019). The Implicit Commitment of Arithmetical Theories and Its Semantic Core. *Erkenntnis*, *84*(4), 913–37.
Parikh, Rohit (1971). Existence and Feasibility in Arithmetic. *Journal of Symbolic Logic*, *36*(3), 494–508.
Shapiro, Stewart (2009). We Hold These Truths to Be Self-Evident: But What do We Mean by That? *Review of Symbolic Logic*, *2*(1), 175–207.
Tait, William (1981). Finitism. *Journal of Philosophy*, *78*(9), 524–46.
Tennant, Neil (2002). Deflationism and the Gödel Phenomena. *Mind*, *111*(443), 551– 82.
Titelbaum, Mike (2015). Rationality's Fixed Point. *Oxford Studies in Epistemology*, *5*, 253–94.
van Fraassen, Bas (1980). *The Scientific Image*. Clarendon Press.
Weatherson, Brian (2019). *Normative Externalism*. Oxford University Press.
Wright, Crispin (2004). Warrant for Nothing (and Foundations for Free)? *The Aristotelian Society Supplementary Volume*, *78*(1), 167–212.