MALAMENT-HOGARTH MACHINES AND TAIT'S AXIOMATIC CONCEPTION OF MATHEMATICS

Abstract. In this paper I will argue that Tait's axiomatic conception of mathematics implies that it is in principle impossible to be justified in believing a mathematical statement without being justified in believing that statement is provable. I will then show that there are possible courses of experience which would justify acceptance of a mathematical statement without justifying belief that this statement is provable.

1. Introduction

In his 2001 paper, *Beyond the axioms: The question of objectivity in mathematics*[11], William Tait advances an axiomatic conception of mathematics on which provability constitutes the sole 'criterion' for mathematical truth. According to this view, proof is ultimately the only source of epistemic justification for mathematical beliefs. As Tait puts it, "the assertion of a mathematical proposition is warranted only by a proof of it"[1].

In this paper, I will argue that the account of mathematics in *Beyond the Axioms* implies that it is impossible to be epistemically justified in believing a mathematical statement without being justified in believing that statement is provable. Against this claim, I will argue that certain physically possible courses of experience (as of dealing with hypercomputers) *would* epistemically justify belief in a mathematical statement without justifying belief that this statement is provable[2].

[1][11]pg.11

[2]I'd like to thank Warren Goldfarb, Peter Koellner, Ned Hall and Peter Gerdes for help fine-tuning my Tait exegesis and much lively debate about the larger philosophical questions at issue in this paper.

2. Tait's Axiomatic Approach to Mathematics

2.1. **The Axiomatic Approach.** At the beginning of *Beyond the Axioms*[11] Tait criticizes forms of realism which make mathematics 'speculative' in the sense that, "even the most elementary computations, deductions and propositions must answer to a reality which we, at best, can only partially comprehend and about which we could be wrong."[3] Instead, he proposes an axiomatic approach to mathematics, on which any consistent mathematical practice we adopt would give meaning to our mathematical expressions in such a way as to ensure the truth of whatever statements this practice instructed us to accept.

This account claims to be a form of realism, in the sense that it takes mathematical statements to be literally true and to stand in no need of paraphrase. However, it denies that our choice of mathematical axioms is answerable to an independent reality which these axioms partly and fallibly describe. As a result, foundational worries about whether our most fundamental axioms might be false are necessarily unfounded. We can acquire particular false mathematical beliefs if, for example, a slip of the pencil leads us to falsely believe that a statement is derivable from our axioms, but it would be impossible for statements which we accept as axioms to themselves be false. The only way our axioms can fail to express truths is if they are inconsistent – which Tait thinks would make our relevant mathematical talk meaningless rather than false[4]. Accordingly, the traditional realist's problem of accounting for human knowledge of mathematics reduces to a problem of explaining our ability to choose consistent axioms.

[3] [11] pg.4

[4] He writes, "if we should discover a contradiction in Peano Arithmetic, say, that would ... undermine the *sense* of existence assertions concerning numbers (and so the sense of their negations, as well)"[11] pg.3

2.2. **Contrast With More Conventional Approaches.** To appreciate how radical Tait's proposal that all consistent mathematical practices would express truths is, it will be important to distinguish the (essentially syntactic) notion of consistency which Tait appears to have in mind in *Beyond the Axioms* from a more demanding notion, coherence, which is relevant to contemporary discussions of structuralism, plenitudinous platonism and other similar philosophical interpretations of mathematics. To do this, we must first clarify what Tait means by proof and mathematical practice.

In Tait's sense, a mathematical practice is a practice of accepting certain premises (the axioms) and certain kinds of inferences within the context of a mathematical argument. Note that, in this sense, our mathematical practice reflects our full judgment of what proofs suffice to establish their conclusions, as opposed to merely establishing that these conclusions can be proved from certain arbitrary axioms. Thus, for example, our mathematical practice (in the sense which Tait has in mind) will plausibly include something like the combination of first order logic and the standard Zermelo-Fraenkel (ZF) axioms of set theory[5]. Also, note that while some might be inclined to call various infinitary objects proofs, Tait's usage in *Beyond the Axioms* makes it clear that he takes proofs be not only finitary but also algorithmically verifiable. For, Tait assumes that Gödel's incompleteness theorem applies to our mathematical practice[6]. Thus it appears that (for the purposes of this paper) Tait takes our mathematical practice to be algorithmically describable, and means to avoid any non-computable notion

[5]It will plausibly also include more than this. For although all propositions which can be derived by applying first order logic to the Zermelo-Fraenkel axioms would be accepted by mainstream mathematicians, there are further statements, such as the arithmetical sentence Con(ZF), which cannot be proved from these axioms yet are taken as genuine items of mathematical knowledge and perfectly acceptable starting points for proofs.
[6]"But, of course, completeness fails and must fail. Nor is the essential incompleteness due simply to Gödels incompleteness theorem."[11] pg. 3.

of consequence or specification of axioms that might evade the assumptions of Gödel's theorem.

With this in mind, I will say that a mathematical practice is *syntactically consistent* if it does not permit one to derive both a statement and its negation[7]. I will also say that a mathematical practice is *coherent* if it is (mathematically) possible to simultaneously satisfy its axioms, and all theorems which its inference rules allow one to prove from these axioms[8]. When the allowed inferences are clear from context, I will speak of theories being syntactically consistent or coherent. Note that the statements we make in our mathematical practice seem to require things that go beyond there first order consequences[9]. We will see that in such circumstances the notions of consistancy and coherence can come apart.

Many philosophers with structuralist, plenitudinous platonist or fictionalist leanings are attracted to the idea that all[10] and only coherent theories are legitimate topics for mathematical investigation. Accordingly, they accept that all coherent mathematical practices would lead mathematicians to express only truths if they were adopted. However, they do not accept Tait's stronger claim that all syntactically consistent practices would lead mathematicians to express only truths, because they think that syntactically consistent theories need not be coherent.

---

[7]i.e. if one cannot derive both a statement and its negation using the premises and inference rules permitted by that practice

[8]For instance, the existence of a model demonstrates the coherence of a theory but, when working in strong logics, it may not be the case that all coherent collections of axioms have a set model e.g. if one can uniquely describe the structure of the sets then this description will not apply to any structure inside the universe of sets

[9]e.g., the statement that every set has a powerset intuitively requires something that isn't true in a countable model of ZFC

[10]In the case of the plenitudinous platonist I am speaking loosely with regard to the claim that all coherent theories are acceptable topics for investigation. For appeal to quantifier restriction, broad limits on abstraction or something more is needed to deal with the fact that not all internally coherent descriptions of mathematical objects are compatible with one another.

These considerations are more than hypothetical, as most most mathematicians and philosophers of mathematics accept non-first-order claims about mathematical subjects (number theory, analysis, set theory etc.) which they take to uniquely describe the intended structure of the numbers [11] – and thus constrain the meaning of their terms in such a way as to give definite truth values to all statements in the language of number theory. Insofar as no algorithmic proof procedure can (correctly) decide all questions of number theory, any proof practice whose axioms included such descriptions of the numbers would have to fix truth values for certain number theoretic statements which it did not allow one to prove or refute.

Accordingly it appears that syntactically consistent extensions of our mathematical practice can nonetheless fail to be coherent. Tait rejects this line of reasoning by denying that we can have any such proof-practice-transcendent grip on the intended structure of the numbers. He denies that there can be any " upon which a proposition, undecided by our present axioms, is nonetheless really true or really false."[12] Thus, he denies that claims like, 'The numbers are as small as possible while satisfying Peano Arithmetic' can take on a meaning which requires the truth (or falsehood) of statements that our mathematical practice doesn't let us prove or refute.

Motivated by roughly Wittgensteinian concerns about manifestability, Tait criticizes various proposals about what such a proof-transcendent understanding of the intended structure of the natural numbers could consist in. He points out that, for example, any definition we give of the intended structure of the natural numbers will itself use terms that need to be

[11]For instance, they believe the numbers are as small as possible while satisfying certain basic principles of arithmetic. However, no consistent collection of first-order axioms can fully express this idea. Any first-order theory that describes the natural numbers will also be satisfied by some non-standard model including infinite 'numbers.' Thus, it would seem that our real axioms for the numbers go beyond what is first-order expressible, in ruling out these spurious infinitary 'numbers.

[12][11] pg. 13

antecedently understood, that any mental picture we associate with these words will need interpretation and that facts about the neural mechanisms which underly our actual use of mathematical terms cannot ground a distinction between correct use and psychologically natural and ingrained but incorrect use.

This skepticism about claims to a proof transcendent grip on logical and mathematical notions puts Tait in a position to accept the shocking claim that any syntactically consistent theory, including any syntactically consistent extension of the axioms which we allow to figure in our reasoning about the numbers, could express a truth if we chose to extend our mathematical practice in the appropriate way. In cases where a claim about the structure of the numbers is not provable or refutable from axioms, he holds that we are *free to stipulate either answer*. So, for example, Tait must accept that we are free to stipulate the truth or falsity of any number theoretic statements which are genuinely independent of our mathematical practice. Indeed, in the case of the Continuum Hypothesis (a set theoretic statement which can be shown to be independent of the generally accepted ZFC axioms of set theory) he explicitly says that, "Until we determine it, CH [the contiuum hypothesis] is ... indeterminate"[13] and there may be equally good directions in which our conception of set could develop which would require us to adopt axioms which imply either the continuum hypothesis or its negation.

2.3. **Independence and the Axiomatic Conception of Mathematics.**

One might think that any theory which takes mathematical practices to play as direct a role in determining mathematical truth as Tait's does would face problems with Gödelian incompleteness. However, Tait, while accepting that the premises of the Incompleteness theorem apply to the system of

[13] [11] pg.13

axioms and inference rules which constitute human proof practices, evades the obvious problems.

For example, allowing that some mathematical statements will be neither provable nor refutable from axioms we are inclined to accept does not prevent Tait from making sense of mathematicians' use of the law of the excluded middle. Insofar as all statements of the form $\phi \vee \neg\phi$ are provable in classical logic, and classical logic forms part of accepted mathematical proof procedures, these statements will come out to be true on Tait's account - even in cases where neither ϕ nor $\neg\phi$ is provable. Furthermore, Tait suggests that in taking our mathematical practice to ensure the truth of $\phi \vee \neg\phi$ we should take it to ensure the truth of statements of the form 'ϕ is true or ϕ is false' as well. Thus, Tait accepts all the same claims as the standard Platonist about principles derived from the law of the excluded middle.

The difference between Tait's view and more conventional views only emerges when we consider the possibility of introducing new mathematical axioms. As we saw above, platonists, structuralists and some fictionalists will rely on the categoricity of our non-first-order description of the numbers[14] to conclude that our current understanding of the intended structure of the numbers already suffices to determine a definite answer to the question of whether any number theoretic claim is true, whether we know it or not. As a result, they maintain that we are not free to sharpen our mathematical notions by stipulating an arbitrary answer to this question.

In contrast, Tait holds that there are no grounds on which a statement which is independent of our axioms could be right rather than wrong. He denies that we have any proof-transcendent grasp of the intended structure of the numbers which determines a right answer to statements independent

[14] That is, on the fact that these descriptions uniquely determine a mathematical structure.

of our axioms. As a result we are free to stipulate the truth value of independent mathematical statements at will. Thus, although Tait and the more mainstream philosophers above will both assert that exactly one of ϕ and $\neg\phi$ is correct, Tait will go further and say that we are free to stipulate which of these statements is correct.

2.4. **Warrant.** Finally, Tait's axiomatic understanding of mathematics leads him to advance a constraint on mathematical justification which will be the main focus of this paper. He claims that provability is the sole 'criterion' for mathematical truth[15] in the sense that, "the assertion of a mathematical proposition is warranted only by a proof of it"[16]. This claim requires a little unpacking.

One might object that we routinely assert mathematical claims without having access to a proof, for example, in response to looking at a calculator or listening to an expert. However, both of these examples can be accommodated by charitably interpreting Tait to say that proof is the sole source of mathematical knowledge and epistemic justification for believing mathematical claims, in the following sense: one can only by epistemically justified in believing a mathematical claim ϕ by considering a proof ϕ *or* having evidence which justifies the belief that ϕ is provable. Plausibly, looking at a calculator only justifies us in believing a mathematical claim when it also justifies us in believing that this mathematical claim is provable via whatever process of accepted mathematical derivation the calculator is designed to replicate. Similarly, one might think that testimony from someone who claims to have inspected a proof of some mathematical claim ϕ only justifies acceptance of ϕ if it justifies belief that this person has gone through a proof of ϕ or otherwise justifies belief that this claim is provable.

[15] see [11] pg. 4, 8-9
[16] [11] pg. 11

Tait motivates the claim that proof is the sole criterion for mathematical truth by appeal to the ideas about axiom choice noted above. Suppose there were (contra the hypothesis above) some independent source of justified mathematical beliefs, which went beyond proofs and reason to think that a given proposition was provable. Plausibly, any such justification would have to involve (apparent) awareness of some 'grounds' besides provability from the axioms, upon which a mathematical proposition could be true rather than false. But such grounds would also have to be "grounds upon which a proposition, undecided by our present axioms, is nonetheless really true or really false."[17] And this in turn, would seem to provide a sense in which the addition of consistent axioms could be wrong. In particular, if some proof-independent method of mathematical learning could teach us that an independent sentence ϕ was true, this would seem to provide a sense in which adding the negation of ϕ to our axioms could be consistent but wrong.

Now, the above argument that proof is the only source of justification for mathematical claims is a purely a priori one, and makes no appeal to contingent facts. Thus if it succeeds, it establishes not only the absence but the *in principle impossibility* of alternative sources of mathematical justification. I will now attack this claim by arguing that certain physically possible courses of experience (experiences as of performing experiments with hypercomputers) would justify belief in a mathematical statement without justifying belief that this statement was provable.

3. Malament-Hogarth Machines and Independent Sentences

My argument begins with the consideration of Π^0_1 sentences, that is, simple statements in the language of arithmetic which are writable in the form

[17][11] pg. 13

$(\forall n)F(n)$ where $F(n)$ contains only bounded quantifiers[18]. The fact that (as noted above) the premises of Gödel's theorem apply to our mathematical proof practice, ensures that there will be some Π^0_1 sentence which this proof practice does not allow us to prove or refute[9][19].

I will argue that experience with certain physical hyper-computers could justify belief in the truth of such independent sentences. David Malament and Mark Hogarth have pointed out that certain solutions to the equations of general relativity would allow a person (the operator) and a computer to take different paths through space-time in such a way that the following strange thing happens: no matter how long it takes for the computer to signal its result, the operator will receive that signal within a bounded amount of time, e.g., the operator would receive the result within a single day no matter how many steps occur in the computation[7].

A person exploiting this set up would be able to 'compute' things that a Turing machine cannot. Assuming limitations on memory, power and reliability can be overcome, such a person could seek evidence for the truth of an arbitrary Π^0_1 sentence ϕ by programming the computer to check all of ϕ's instances. For example, if they were interested in in the Goldbach conjecture, they would first program a computer to check that 4 is the sum of two primes, 6 is the some of two primes and so on, signaling back if it ever finds an even number that isn't the sum of two primes. Then they would then launch this computer on a path such that any signal sent by the computer would reach them within a day. If they do not receive a

[18]Thus, for example, the Goldbach conjecture states that every even number greater than 2 is writeable as the sum of two primes. This qualifies as a Π^0_1 sentence because it requires that $\forall n$ $n = 2$ or n is odd or $\exists x \leq n \exists y \leq n$ and x is prime and y is prime and $x + y = n$, where the property of being prime is itself expressible using only bound quantifiers.

[19]The incompleteness theorem applies to any collection of mathematical statements, such as those which could be derived using a particular mathematical practice, which is syntactically consistent, algorithmically enumerable, and sufficiently powerful to capture certain basic facts of number theory. It tells us that, any such collection will fail to include both some Π^0_1 sentence and its negation.

signal within this day (and if there is no stage at which the computer fails to transition as required by its program) then we can infer the Goldbach conjecture is true. Call the whole system consisting of the computer, means to launch the computer through a suitable region of space-time and the signaling mechanism a Malament-Hogarth (or MH) machine. MH machines can be made to test every Π^0_1 sentence using the strategy just described[20].

Discussion of Malament-Hogarth machines in the literature has centered on questions about the physical possibility of the space-time structure needed for an MH machine as well as whether MH machines would count as computers. For my purposes, however, all that will matter is that some course of experience could justify the belief that one was dealing with such a machine. I will argue that, contra Tait's claim, experiences as of dealing with an MH machine can epistemically justify belief in an independent Π^0_1 sentence without justifying belief that this sentence is provable[21].

---

[20]The standard definition of an MH machine in the literature requires that the machine only go through stages corresponding to natural numbers, i.e. it computer accepts the Π^0_1 sentence iff no actual integer provides a counterexample. However, one thing that's at issue here is whether we can think thoughts which distinguish a unique structure ω from various 'nonstandard models'. Therefore, I adopt a less restrictive notion which merely requires an MH machine to include a computer which checks 0 and then (provided no counterexample has been found) checks the successor of any stage it checks.

[21]The reader may wonder why I don't take the much simpler route of simply arguing that our continued failure to find a contradiction while working with certain proof systems (e.g., ZFC set theory) gives us reason to accept the sentence expressing the arithmetical consistency sentences for these practices.

However, there is a significant line of worry in the literature about whether merely using a system without encountering a contradiction can give us reason to believe that that system is consistent, or whether the reason it gives us can be sufficient to let us qualify as having knowledge. Following Frege, some philosophers have argued the numbers differ from one another so radically that "in the absence of proof, we should not expect numbers (in general) to share any interesting properties."[22] and hence that dealings with any number of finite cases where some number has failed to code a proof of 0=1 in a given proof system S can never provide us with any justification at all for the belief that some (untried) number fails to code a proof of contradiction in S. Less radically, it is sometimes argued that dealings with particular cases always provide us with a biased sample - with knowledge of what holds for small numbers and short proofs, and that such knowledge provides no basis for justified generalization to the claim that all numbers have a certain property or that no larger proof is possible.[10]

4. Believing in MH machines

It seems fairly clear that experiences which justified the belief that one was dealing with a genuine MH machine could thereby suffice to justify believing the mathematical result indicated by the that machine. After all dealings with a calculator can justify mathematical beliefs, and a MH machine merely performs infinitely many such computations. All that remains to show is that some experience could justify the belief that one was dealing with an MH machine. To this end, let us consider, in detail, what is required to justify such a belief.

First, one must believe that one's universe has the right kind of space-time structure. Could any experience justify this belief? I take the existence of actual evidence-heavy debates about whether the laws of physics are compatible with the needed space-time features to suggest that it can [4, 3]. Moreover, the history of physics gives a straightforward picture of the kind of evidence which would suffice to justify such a belief.

Second, one must believe that a computer (or machine if you prefer) can be constructed with access to sufficient memory and power. Turing machines, as mathematical abstracta, are allowed to use an unbounded amount of memory and operate for indefinitely many stages. However, the ordinary computers that we build only have access to a finite amount of memory and power. These facts present a problem since an unbounded amount of memory and indefinitely many operations are needed to check arbitrary Π^0_1 sentences.

A traditional answer to worries about energy in discussions of the physical possibility of (memory-limited versions of) MH machines draws on the fact that the computer doing the computations travels infinitely far: one

I avoid this obstacle by providing an apparent counterexample to Tait's thesis which does not depend on the claim that a history of safe use of a mathematical theory can provide justification sufficient for knowledge of the claim that that theory is consistent.

could build the computer to harvest energy as it travels[1]. In a universe like that theorized by some early twentieth century astronomers, matter appears spontaneously in empty space at a certain rate. This would provide a guaranteed source of energy which the machine could harvest on its journey. Similarly, the computer could draw on this energy (converting to matter as needed) to construct further memory cells as necessary.

Alternately, in a 'gunky' universe which allows for complexity at an arbitrarily small scale there is a more elegant solution to concerns about power and memory. With the same strategy used by the electronics industry (shrink components to pack more functionality into a smaller, and hence more energy efficient, package) the computer could continually replace itself by a more efficient copy using less energy but with more memory. Provided one increases efficiency at a sufficient rate the total energy needed would be finite.

Third, one needs to believe that the traveling computer one has constructed is sufficiently accurate to perform as designed throughout its journey. Suppose that the computer being launched has some constant (independent) probability $\epsilon > 0$ of making an error at every given stage in the computation. Then probability that the computer makes it through n stages without failure is $(1-\epsilon)^n$. Thus, the probability of completely correct performance goes to 0 as n goes to infinity. So it would seem that the probability that a Malament-Hogarth machine has worked as intended when verifying a Π^0_1 sentence should be 0. For a related discussion see [2].

However, we can apply the same strategy we used above to make our MH machine as reliable as desired. As well as increasing in capability over time we engineer our computer to improve its reliability as well. Well known techniques in circuit design redundancy and error correction can be used to arbitrarily reduce the probability of an error. By increasing reliability sufficiently quickly we can make the overall chance of machine malfunction arbitrarily small. In this way, dealings with an MH machine can produce degrees of justification which come arbitrarily close to one's justification for believing one's overall physical theory is correct.

Alternately, if you don't like the idea of engineering an indefinitely self-improving computer as above, one can also circumvent worries about energy and error-rates by appeal to possible physical laws which directly constrain the behavior of physical objects that can be used to build a computer. One can imagine an MH machine computer whose basic components were fundamental particles whose behavior was completely determined by fundamental physical laws. In this way there seem to be conceivable and elegant systems of fundamental physical laws which would imply the perfect functioning of various simple building blocks for an MH machine, and hence the perfect functioning of the MH machine as a whole. As no truly bizarre physical laws are required for this scenario, we have every reason to believe that some course of experience would justify concluding that one had built an MH machine in such a manner.

Of course, neither of the accounts above eliminate the possibility of systemic error, e.g, the epistemic possibility that the general physical theory which you used to calculate error rates when designing the MH machine could be incorrect. I have provided some reason for thinking that one's justification for believing one had launched a genuine MH machine could approach one's justification for accepting our most certain physical theories. But, if mathematical knowledge required a substantially different and higher standard of justification than physical knowledge, it might seem that this degree of justification could never suffice to underwrite mathematical knowledge. As a result one might worry that dealings with an MH machine could never provide *sufficient* justification to ground mathematical knowledge.

Note, however, that we cannot say that mathematical knowledge requires certainty on pain of ruling out mathematical knowledge by testimony. And it would seem that beliefs about the physical structure of space and the components in an MH machine can acquire justification on par with one's justification for believing a credible mathematical witness. Thus, it would seem that experiences can provide us with at least the degree of justification which suffices to grant us mathematical knowledge in cases where we learn we new mathematical truths by accepting credible mathematical testimony.

In light of these considerations, I conclude that a suitable course of experience could justify someone in believing that they had built a working MH machine and thereby in believing any Π^0_1 sentence which this machine appeared to verify.

5. Responses

Now let us return to Tait. I have argued above that dealings with a Malament-Hogarth machine could justify asserting Π^0_1 sentences, even when those sentences are unprovable. This is in direct conflict with the central tenets of Tait's account of mathematics which, as we saw, require that (evidence for) proof be the only possible source of justification for asserting a mathematical claim.

I will conclude by considering some responses to the line of argument above.

5.1. **Epistemic vs. Pragmatic Justification.** First, defenders of Tait might resist my claim that experiences as of dealing with MH machines can provide *epistemic* reason for accepting Π^0_1 sentences. Perhaps such experiences don't show that the relevant Π^0_1 statement currently expresses a truth, but only provide pragmatic reasons to study 'systems of number theory' where this sentence is accepted as an axiom.

Although immediately unintuitive, this way of understanding the relationship between observations as of MH machines and number theoretic facts can be motivated somewhat by considering an analogy to geometry. The great success of Euclidean geometry in describing space[23] made it practically useful and convenient for the Greeks to study mathematical systems which included the parallel postulate. It is plausible that adopting (or rejecting) the parallel postulate for reasons like these does not involve learning that any antecedently understood mathematical statement expresses a truth. Instead, it involves pragmatically choosing to study a given mathematical system, because facts about this system appear to have a certain desirable relationship to facts about the external world. At first glance the role which I have argued that MH machines can play in justifying number-theoretic beliefs can seem similar to the role of physical applications in motivating the choice of axioms for geometry.

However, I think this kind of defense is ultimately quite difficult to maintain. Accepting it would require us to reject a certain aspect of mainstream mathematical practice (or at least, mainstream mathematical belief-revision dispositions) as irrational. If experience just makes it rational to study systems in which a certain axiom is true, we ought not to conclude that a mathematical statement is false in response to failures of these applications, but only (at most) that other mathematical systems deserve attention as well. And in the case of geometry this is exactly what happened. When experience motivated the study of non-euclidean geometry we did not say euclidean geometry was wrong but only that other kinds of geometry were worth studying as well. In contrast, in the case of arithmetic our dispositions to revise beliefs are quite different. Learning that an MH machine 'verified' a Π^0_1 sentence would make people say that they were wrong to ever believe

[23]At least near earth for the kind of low-tech uses that can ignore relativistic effects

its negation. Thus, in contrast with the geometrical case such an experience would lead us to dismiss our previous beliefs as wrong about the numbers rather than right about some other system.

5.2. **Exception for Π^0_1 Sentences?** Second, some readers may feel that the counterexample presented in this paper is not very deep, because it depends on exploiting special features of the simplest possible kind of independent statements - independent Π^0_1 sentences. Thus, one might think that although I have presented a genuine counterexample to Tait's view as stated, my objection can easily be handled by a simple modification of Tait's view which takes Π^0_1 sentences to be a special exception to his general claim that there can be no "grounds upon which a proposition, undecided by our present axioms, is nonetheless really true or really false." I will now argue that no such quick fix solution can be given, without giving up central parts of Tait's view.

Π^0_1 sentences are special in the following sense: if a Π^0_1 sentence is consistent then it is true[24]. Because Π^0_1 sentences have the form $\forall x \phi(x)$ where $\phi(x)$ is quantifier free, if such a sentence is false then there is some particular number n which constitutes a counterexample. Since ϕ lacks any unbounded quantifiers, this latter statement will be provable via basic arithmetic rules. As a result, uncontroversial set theoretic reasoning about the numbers allows us to prove that any Π^0_1 sentence of arithmetic which is independent from our overall theory (or even just the part of it summarized by the Peano Axioms) must be true[8].

More traditionally realist readers (like myself) will be inclined to think that these considerations point out a clear ground upon which a mathematical sentence which is independent of our axioms can be right rather

[24] i.e. if one can add a Π^0_1 sentence to any proof practice which (like ours) contains the Peano Axioms and first order logic, then this sentence expresses a truth [finalcheck!]

than wrong. Objective facts about derivability in formal systems like PA combine with our expectations about the relationships between arithmetical sentences and derivability to ensure that independent Π^0_1 sentences are "really true" despite our inability to prove this fact.

However, I do not think that Tait could accept the above argument. The argument above crucially turns on taking our beliefs about the numbers to latch on to objective proof-transcendent facts about derivability in formal systems, and make the truth or falsehood of undecidable sentences reflect these objective facts. Allowing this immediately threatens Tait's core motivating idea that our choice of axioms in mathematics is not a matter of speculating about some independent partly understood subject matter.

If one allows that there are such objective and determinate proof-transcendent facts about derivability in formal systems, and that the meanings of our words can latch on to these facts then these facts about provability would seem to constitute a legitimate subject matter for investigation. Which sentences are provable from which formal systems? Thus, there would seem to be a portion of mathematics at least (the study of derivability in formal systems) where "even the most elementary computations, deductions and propositions" *are* answerable to "a reality which we, at best, can only partially comprehend and about which we could be wrong."[25] Thus, I think Tait must deny that our talk about consistency and derivability latches on to any such proof transcendent facts about derivability[26].

[25][11] pg.4

[26]Admittedly, adopting this line of response raises serious problems of its own. For example, what sense are we to make of Tait's own talk of consistency when saying that, e.g., inconsistency debars an axiom system from giving meaning to our mathematical claims? If claims about consistency are only determined to have a particular truth value by being derived in some axiom system, what axiom system is relevant to Tait's claim? If the relevant axiom system is the total collection of mathematical claims we are inclined to accept, there's a prima facie problem. This system (presumably) cannot prove its own consistency[6]. If there is not a finite demonstrable inconsistency in our axioms, then the question of whether the total collection of axioms that we are inclined to accept determine

Furthermore, it is not clear that Tait can accept the above account of provability for Π^0_1 sentences while maintaining that we are free to stipulate right answers to *any* quantified sentences in the language of arithmetic. If considerations of consistency are sufficient to provide a sense in which all Π^0_1 sentences are "really" right or wrong (even in cases where they are not derivable or refutable) then similar considerations allow us to ground the truth-value of all arithmetic sentences. Thus, in attempting to make a special exception for Π^0_1 sentences we end up extinguishing any role for stipulation in arithmetic and instead force it to answer to an independent .

The key point to note is that there is a direct relationship between whether a sentence with $n+1$ quantifier alternations is true and the facts about what true (or false) statements with n quantifier alternations can be proved from it. In particular, a sentence that begins with a universal quantifier and includes $n+1$ quantifier alternations followed by some formula containing only bounded quantifiers (called Π^0_{n+1} sentence) is true if and only if it adding it as an axiom to Peano Arithmetic does not allow one to prove some false Σ^0_n sentence, (i.e., some statement that begins with an existential quantifier and includes n quantifier alternations followed by a formula with only bounded quantifiers). To see why this is so, consider the example of arbitrary Π^0_2 sentence $\forall x \exists y \psi(x, y)$ where ψ is a formula in the language of arithmetic with only bounded quantifiers. This sentence is false iff $\exists x \forall y \neg\psi(x, y)$ is true, i.e, if there is some number n such that $\forall y \neg\psi(\overbrace{S \circ S \cdots \circ S}^{n}(0), y)$ is true. But, from $\forall x \exists y(x, y)$ one can derive (over Peano arithmetic) the negation of each such instance. Thus, an arbitrary Π^0_2 sentence is false if and only if adding it as an axiom allows you to derive some false Σ^0_1 sentence.

a consistent and hence true mathematical system, or an inconsistent (and hence meaningless) one will turn out to have the same status as independent Π^0_1 sentences. This seems like an odd consequence. It also seems odd that it in stipulating facts about arithmetic we could thereby determine facts about what alternative choices of axioms would have been meaningful.

As a result, if we are only free to stipulate new mathematical axioms in a way that honors our current expectations about the relationship between provability and truth in arithmetic, then if our hand is forced with regard to Σ^0_n sentences it will be forced with regard to Π^0_{n+1} sentences as well. But, noting this fact allows us to inductively show that our lack of any freedom to choose truth values for bounded sentences permeates all the way up, and we are never free to choose how to settle the truth value of any quantified sentences in the language of arithmetic. We can inductively fix mandatory truth values for all such sentences as follows:

- $\Pi^0_0 = \Sigma^0_0$ sentences are true iff they are provable in PA.
- Σ^0_{n+1} sentences are true iff the Π^0_{n+1} sentences which form their negations are false
- Π^0_{n+1} sentences are true iff adding them to Peano Arithmetic as an axiom does not allow you to derive some false instance, i.e., a false Σ^0_n sentence.

Finally, even if Tait could somehow motivate making a special exception for Π^0_1 sentences but not more complex sentences of arithmetic, he would still run into a version of the problem presented by MH machines. Hogarth has demonstrated that variants of the MH machine can be constructed to check arbitrary sentences in the language of number theory [7]. The key idea is to consider a Malament-Hogarth machine which spawns other MH machines. Thus, to check the truth of a $\forall x \exists y \phi(x, y)$ sentence one builds a computer which checks whether $\exists y \phi(1, y)$, by building and launching a standard MH machine computer which is set to look for a y such that $\exists y \phi(1, y)$, and radio back to the main computer if it ever finds one. If the main computer doesn't receive a signal within the relevant interval, it decides that $\neg \exists y \phi(1, y)$ so it has found a counterexample, and it radios back that $\forall x \exists y \phi(x, y)$ must be false. If it does receive a signal from the child computer it decides that

$\exists y\phi(1, y)$ is true and proceeds to check whether $\exists y\phi(2, y)$ and all other instances in the same way, using the property of MH machines to accomplish all this checking in bounded time for the operator. Thus, if we do not hear back from the master computer within the relevant interval we can conclude that the relevant $\forall x\exists y\phi(x, y)$ sentence is true. A similar extension of the machinery allows one to check the truth of all arithmetic sentences. Thus, it would seem that the possibility of getting (partly empirical) justification for believing arithmetical statements which are independent of all axioms we are inclined to accept is not limited to Π^0_1 sentences.

6. Conclusion

In this paper I have argued that experiences as of creating a Malament-Hogarth machine could provide epistemic justification for believing independent mathematical sentences, justification which does not appeal to any reason to think these statements are provable from axioms we are disposed to accept.

If this is correct it constitutes a counterexample to Tait's claim that proof is the only possible source of warrant for asserting a mathematical claim. And insofar as we have seen that central features of Tait's axiomatic understanding of mathematics lead him to the conclusion that only proof can justify mathematical claims, reason to deny this epistemic claim casts doubt on this theory as a whole.

Our discussion of MH machines also provides reasons for doubting projectivist and pragmatist approaches to truth in arithmetic. We have seen in the pages above that we expect the right answers to questions about the numbers to be reflected by certain constraints on how it is metaphysically possible for infinite physical systems like MH machines to behave. Insofar as this is the case, we are not free to pragmatically stipulate right answers.

This result is interesting in itself and has implications for many philosophical interpretations of mathematics, though none so dramatic as the implications it has for Tait.

References

[1] Tim Button. Sad computers and two versions of the church–turing thesis. *The British Journal for the Philosophy of Science*, 60(4):765–792, 2009.

[2] J. Earman and J. Norton. *Infinite pains: the trouble with supertasks*, volume 11, page 271. Blackwell Publishers, Cambridge, Mass., 1996.

[3] J. Earman and J.D. Norton. Forever is a day: supertasks in Pitowsky and Malament-Hogarth spacetimes. *Philosophy of Science*, pages 22–42, 1993.

[4] G. Etesi and I. Németi. Non-turing computations via Malament–Hogarth space-times. *International Journal of Theoretical Physics*, 41(2):341–370, 2002.

[5] Gottleib Frege. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number* Northwestern University Press, 1980

[6] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter systeme I. *Monatshefte für Mathematik*, 38(1):173–198, 1931.

[7] Mark Hogarth. Deciding arithmetic using SAD computers. *The British Journal for the philosophy of Science*, 55(4):681–691, 2004.

[8] Richard Kaye. *Models of Peano Arithmetic*, volume 15 of *Oxford Logic Guides*. Oxford University Press, New York, 1991. Oxford Science Publications.

[9] B. Rosser. Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, 1(3):87–91, 1936.

[10] Alan Baker. Non-deductive methods in mathematics. *Stanford Encyclopedia of Philosophy (Fall 2013 Edition)*, Edward N. Zalta (ed.), URL = ¡http://plato.stanford.edu/archives/win2012/entries/davidson/¿.

[11] W.W. Tait. Beyond the axioms: the question of objectivity in mathematics. *Philosophia Mathematica*, 9(1):21–36, 2001.