MH MACHINES AND EMPIRICAL JUSTIFICATION IN MATHEMATICS

1. Introduction

Can a person's mathematical beliefs ever depend on sensory experience for their justification? At first glance, one might think this is impossible. However, as Kripke pointed out, experience with calculators certainly seems to play a role in justifying some of our mathematical beliefs[Kripke, 1972]. Experience also seems to play a role in justifying mathematical beliefs formed on the basis of testimony[1].

In view of these examples, one may be tempted to allow that mathematical beliefs can depend on experience for their justification but insist that a priori argument is still the *fundamental* source of mathematical knowledge in the following sense: any experience which justifies believing a mathematical claim ϕ does so by way of justifying the claim that it would be possible to give an a priori proof of ϕ.

In this paper I will argue that the weakened proposal above still fails. There are some possible courses of experience which would justify believing a mathematical proposition without justifying belief that the relevant proposition is provable. In particular, I will argue that someone could empirically learn that the laws of their universe allowed the construction of a special kind of physical system which I will refer to as a Malament-Hogarth or MH machine. Experiences as of using such a machine could justify believing a mathematical claim ϕ *without* providing any reason to think that ϕ is provable a priori. I will conclude by noting that if this claim is correct,

[1]However, this claim has been seriously disputed by [Burge, 1993] among others.

it raises a problem for certain contemporary philosophies of mathematics which try to ground the truth of mathematical claims in the existence or constructability of a suitable proof.

Interest in Malament-Hogarth machines in the literature has centered on the question of whether they count as computers. No Turing machine can decide the truth value of arbitrary Π^0_1 sentences. Thus, if what the Malament-Hogarth machine does counts as *computation* then the Church-Turing thesis (that computability is Turing computability) is wrong. There has also been some debate about whether the actual laws of our universe would allow for the construction of a Malament-Hogarth machine. However, for my purposes it will not matter whether the physical set up described above qualifies as a computer, or whether such a set up could be created in the actual world. My argument only depends on the claim that it would be metaphysically possible to have experiences which would justify the belief that one has built a Malament-Hogarth machine (or some similar physical system).

2. Malament-Hogarth Machines

Let me begin by introducing the idea of a Malament-Hogarth machine. It has been pointed out by Mark Hogarth and others that General Relativity allows for the existence of space-times in which an observer (person) and a computer can take different paths in such a way that the following strange thing happens: a signal sent after an arbitrary interval of time in the computer's reference frame is guaranteed to reach the observer within a bounded interval of time in the observer's reference frame. No matter how long the computer delays sending a signal, the observer will receive this signal within a certain five minute interval[2].

[2] [Hogarth, 2004]

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 would be able to determine the truth of any Π^0_1 sentence, a claim about the natural numbers of the form $(\forall n)F(n)$ with $F(n)$ a computably checkable property of n. Thus, for example, they would able to determine the truth of the famous unsolved conjecture that every even integer greater than 2 is the sum of two primes.

To illustrate how a MH machine can do this, consider the specific example of verifying the Goldbach conjecture. To verify the conjecture we need to check that each number of the form $2n$ is the sum of two primes. Verifying that some particular number n has this property is a finite process: one only needs to check that each of the finitely many pairs of primes below $2n$ sum to $2n$.

We can check the general claim that all numbers have this property by programming a Turing-computer to march through each natural number of the form $2n$ in turn, and signal to us if and only if it finds one that cannot be written as the sum of a pair of primes. Thus, normally we would expect to receive some signal from the machine at some (unboundedly large) time in the future if and only if the Goldbach conjecture is false.

Now, however by launching the machine on an appropriate path through the MH space-time, our operator puts herself in a position to 'see to the end' of this infinite process of checking, and thus learn the truthvalue of the Goldbach conjecture. All the signals from these infinitely many different stages of computation will reach her within a single 5 minute window. Thus, at the end of this period she will have received a signal if only if the Goldbach conjecture is false.

Exactly the same strategy works for checking the truth of all other Π^0_1 sentences. Indeed, we can program a single MH machine to 'decide' the truthvalue of any specified Π^0_1 sentence, in the following sense. Given any input n which is the Gödel code for some Π^0_1 sentence, the MH machine's receiving apparatus will wait 5 minutes and then (after this finite lapse of time) display a 1 if the sentence is true and a 0 if the sentence is false.

In the following sections I will argue for two claims. First, I will argue that experience could justify a person in believing a Π^0_1 sentence on the basis of seeming to have built something like a Malament-Hogarth machine which verifies this claim. Second, I will argue that a person in this situation could, nonetheless, fail to have any justification for thinking the relevant Π^0_1 sentence was a priori provable.

3. MH Machine Experience and Mathematical Facts

The arguments above amply demonstrate that if someone was justified in believing they had used something like an MH machine to test a mathematical claim they would have justification for believing the mathematical claim. This leaves us with the question of whether someone could be justified in believing they have used such a machine.

I will argue that, whether or not MH machines are physically possible in our world, reasonable tweaks to the laws of physics and the experimental results that follow would justify belief in such machines. Let us consider piece by piece what taking oneself to have built an MH machine involves.

First, there is the question of whether one's universe has the right structure. Could any experience justify one in believing that one inhabited an MH spacetime, or some other universe where the structure of space and time allowed for the kind of signaling described above. I take the existence of actual evidence-heavy debates in physics about whether the best laws of physics as currently known are compatible with our being in a Malament-Hogarth space time to suggest that it can. And, more generally, I take the history of physics to provide examples of the kind of evidence which might rationally convince someone that space-time had a certain structure.

Second, there are questions about memory and sufficient energy to power indefinite computations. Turing machines, as mathematical abstracta, are allowed to use an unbounded amount of memory. But of course, the ordinary computers that we build only have access to some finite amount of memory. Now, a traditional answer to worries about energy in discussions of the *physical* possibility of MH machines draws on the fact that the plain computer travels infinitely far: one could build it to harvest energy from interstellar gas or passing stars. If we are willing to suppose that certain experiments had turned out differently instead of justifying the big bang they would have justified steady state theories that propose matter appears spontaneously in empty space. A situation that might allow for a guaranteed source of energy/matter the machine could harvest on it's journey. By accumulating such energy/matter the machine could continuously construct extra memory and continue to power itself[3].

Thirdly, and most seriously, there are questions about the durability and reliability of the computer that gets launched. No matter how minute any fixed independent probability of error during each stage of the computation would yield a probability of 0 that the machine operated as intended during an infinite sequence of computations. This difficulty, however, is solvable by the same mechanism that allows the machine to construct additional memory and harvest additional power. It is a well known technique in circuit design that redundancy and error correction can serve to arbitrarily

[3]In a universe which allows for a machine of infinite size, this could be done by simply adding to the overall size of the computer. In a 'gunky' universe which is finite in size but allows for complexity at an arbitrarily small scale, this could be done by shrinking the size of memory cells. Imagine, for example, a computer which pauses after every computation to build a fine-grained replica of itself with the same total size but increased storage capacity.

reduce the probability of error. By designing the machine to continually improve the redundancy and error checking mechanisms during it's voyage the total probability of failure for the entire infinite computation could be guaranteed to be quite large[4]

This concern can also be given a more general answer, if we are willing to consider (the possibility of evidence associated with) universes whose basic laws are radically different from our own. It could look like a simple consequence of fundamental physical laws (the kind one would rationally expect to be preserved forever) that the machine in question behaved like this. We might have to imagine evidence that, contrary to what's actually the case, the universe allowed something like perpetual motion and that energy could be created. However, with regard to durability, it seems conceivable that one's best theory of the behavior of some kind of fundamental particle would be that, e.g. it emits a pulse of light every second, or it always behaves like an OR gate with respect to other particles associated with it in a certain way, and that one could use such particles to build the relevant kind of enduring computer.

In light of these considerations, I conclude that a possible course of experience could justify someone in believing that they had built a working MH machine[5] and thereby in believing an arbitrary Π^0_1 sentence which the machine seemed to verify.

4. MH Machine Experience and Provability

Now let us turn to the question of whether the above-mentioned empirical justification for believing a Π^0_1 sentence would also justify believing the in

[4]All that is necessary is that the probability density function for failure with respect to time approaches 0 sufficiently fast. Thanks to REMOVED FOR ANONYMITY for this point.

[5]For convenience, I will understand the term 'Malament-Hogarth' machine loosely enough to encompass the kinds of hyper-computers exploiting infinitely large or gunky universes in the paragraph above.

an a priori justification of the sentence. I will argue that, in contrast with the familiar cases of empirical justification via testimony or a calculator, learning a Π^0_1 truth by way of a MH machine gives one little or no reason to think that the relevant claim is provable a priori.

Recall that a Π^0_1 sentence ϕ says that every number has a certain recursively checkable property F. Thus, given sufficient time we can check with pen and paper that, say, 3 has the property F. Similarly, given sufficient time and memory an ordinary Turing-complete computer can check that the first billion numbers have property F. In contrast ϕ itself makes a claim about infinitely many different numbers, so no finite amount of direct checking will suffice. At any finite time, all we know is that we haven't come up with a counter-example yet.

What normally allows us to learn true Π^0_1 sentences is not direct checking, but rather access to a proof which takes an intelligent short cut. Using principles like mathematical induction we can establish that some property F holds of all numbers by way of a proof which only involves finitely many steps. However, it is far from clear such an intelligent shortcut will always be available for every true statement of number theory ϕ.

In fact, there are good reasons to think that there must be some unprovable Π^0_1 sentences. The Incompleteness Theorem tells us that no recursively axiomatizable system of reasoning about arithmetic (of sufficient strength to prove basic arithmetic facts) can prove all (intuitively) true Π^0_1 sentences. In particular, it shows that every such system S won't be able to prove a Π^0_1 sentence which says, in effect, 'every number fails to code for a proof of $0 = 1$ in system S'. Thus *if the class of good a priori arguments is recursively axiomatizable*, then there will be a true Π^0_1 sentence a priori reasoning cannot lead one to. Thus, there will be a true Π^0_1 sentence, which cannot be known a priori.

More generally, if there is even one Π^0_1 truth such which we don't have a priori reason to believe is provable we get the result that evidence for mathematical truth can outrun evidence for provability by appeal to the appropriate use of an MH machine.

For these reasons, I claim that experiences with an MH machine which would be sufficient to justify belief in a Π^0_1 sentence would not be sufficient to justify belief that the relevant Π^0_1 sentence was provable. We learn general statements of arithmetic a priori by putting together certain ground truths and methods of reasoning about the numbers, in such a way as to ensure the truth of the general statement without separately checking each of its instances. To ensure our a priori justification of arithmetic sentences doesn't go astray we are necessarily limited to only those sentences provable from 'obvious' starting points. A Malament-Hogarth machine, in contrast, works by brute force: directly checking each instance of the sentence in question, and then communicating results back to us within a finite window of time. The results which an MH machine delivers are just directly sensitive to whether the generalization it has been set to check is true; the machine is not sensitive to whether *the kinds of mathematical premises and inference rules that can figure in proofs* can be put together in such a way as to establish the generalization in finitely many steps[6]. Thus we seem to have, as desired, a case where experience justifies a mathematical claim, without justifying the belief that this claim can be given an a priori proof.

[6]One could get around this problem by allowing that an infinite series of inferences, like those involved in the MH machine's brute force checking would count as a proof. Note, however, that in order to learn from such a proof a person would have to not only live infinitely long, but to live through some series of temporal points with the structure of $\omega + 1$. There would have to be some point strictly after each of the computations that checks $F(n)$ from which you could notice that you had finished checking each instance and conclude that the general statement $\forall x F(x)$ was true.

5. Consequences

If my argument in the pages above works, it shows that experience can justify belief in a mathematical proposition ϕ without justifying belief that ϕ is provable. I will now conclude by briefly discussing what consequences can be drawn from this surprising fact.

First I want to stress that my conclusion, that experience can figure in a certain direct fashion in the justification for mathematical beliefs, should not be confused with the stronger empiricist proposal that mathematical truths truths can be justified *purely* on the basis of experience, or on the basis of experience plus 'analytic truths'. The fact that experience plays a role in the justification of mathematical beliefs formed via MH machines does not preclude the possibility that some kind of substantive a priori insight into mathematics (of whatever type the empiricist wants to avoid) also plays a role.

In contrast, I think considerations about Malament-Hogarth machines have real philosophical payoff with regard to views like William Tait's which try to ground the truth of a mathematical claim in the fact that it can be proved. In 'Truth and Proof: the Platonism of Mathematics' Tait advocates a form of realism on which the truth of mathematical statements is closely tied to the existence (or possibility of constructing) proofs. He writes, "A proposition A is true when there is an object of type A, and a proof of [that proposition] is the construction of an object of type A"[Tait, 1986].

Positing this kind of close connection between truth and provability allows Tait to give a clever response to the ancient problem of access to mathematical objects: how can the human mathematical practice of giving proofs be such a reliable guide to the facts about abstract objects like sets and numbers which mathematical claims appear to describe? On Tait's view no access problem arises because our cannons of mathematical proof do not need to reflect facts about some independent mathematical objects. Instead, the existence (or perhaps constructability) of a proof of A *just is* what it takes for A to be true. Thus it turns out to be metaphysically necessary that whenever mathematicians succeed in producing something that satisfies our norms for being a proof of A then A expresses a truth.

Tait's proposal also allows him to capture standard mathematics' use of the law of the excluded middle. Since standard mathematical practice allows the use of the law of the excluded middle, each statement of the form $\psi \vee \neg\psi$ will indeed have a proof. In particular, every such statement will have a one line proof. As a result each instance of the law of the excluded middle will come out to be true on Tait's view, even in cases where, as per the incompleteness theorem, neither ψ nor $\neg\psi$ is provable.

However, if what I have been arguing in this paper is correct, then there is a portion of standard mathematics (or at least, standard dispositions to revise mathematical beliefs) which Tait's view and other similar views cannot capture. We have seen that courses of experience which do not justify belief that a mathematical claim is provable can nonetheless justify the acceptance of that mathematical claim. But it is hard to see how such a view could make sense of this acceptance.

6. CONCLUSION

In this paper I have argued that certain epistemically possible courses of experience could justify a mathematical claim, without justifying believe that the relevant claim is provable. Experience as of performing certain experiments with a Malament-Hogarth machine could justify acceptance of a mathematical claim ϕ without providing any reason to think that ϕ is provable. This example raises a problem for certain philosophies of mathematics which closely tie mathematical truth (or assertability) to provability.

References

[Burge, 1993] Burge, T. (1993). Content preservation. *Philosophical Review*, 102:457–488.

[Hogarth, 2004] Hogarth, M. (2004). Deciding arithmetic using sad computers. *British Journal of the Philosophy of Science*.

[Kripke, 1972] Kripke, S. (1972). *Naming and Necessity*. Blackwell.

[Tait, 1986] Tait, W. (1986). Truth and proof: the platonism of mathematics. *Synthese86*, 69(3).