What is the problem of mathematical knowledge?

Michael Potter

19 December 2006

This is a book about the problem of mathematical knowledge. So it might be useful to try to get clear about what the problem is. Why, in other words, do philosophers take there to be a particular problem, worth devoting a whole book to, about *mathematical* knowledge?

One thing I should say straightaway is that nothing I shall say is at all deep. Another is that my aim here is modest. All I want to do is to try to get a bit clearer about what the problem is, not to propose a solution.

1 What is distinctive about the mathematical case?

With the preparatory throat-clearing out of the way, the first thing I think we need to get clear about is how the problem of mathematical knowledge differs from the problem of logical knowledge. One of the great failed projects in the history of the philosophy of mathematics was logicism, which aimed to show that mathematical knowledge was just a species of logical knowledge. If that project had succeeded, it would thereby have shown that there is no problem of mathematical knowledge, of how we come to know that mathematical theorems are true, distinct from the problem of logical knowledge, i.e. of how we come to know that arguments are logically valid.

But logicism did fail, which is why there is still a problem of mathematical knowledge. What I want to do here, therefore, is to explore what this problem amounts to. I want to get clear, that is to say, about what *more* there is to the problem of mathematical knowledge than there is to that of logical knowledge.

By formulating the question in this way I do not want to be taken as claiming that there is not a problem of logical knowledge. There is. In the *Tractatus* Wittgenstein tried to advance the beautifully simple idea that there is no problem of logical knowledge because in logic there is, in a sense, nothing to be known. This, like logicism, would have been nice if it had worked. But it did not, which is why there is still a problem of logical knowledge.

Notice, though, that there was, nonetheless, something right about Wittgenstein's idea. There is little doubt that he under-estimated the complexity of logic. He thought that it is trivial, and in that he was wrong. But his reason for thinking it is trivial, namely that it says nothing about the world, surely has something right about it. It is a familiar experience of anyone who has taught modal logic that as soon as we mention ‘possible worlds’, some students will immediately assume that this is meant as a contrast with ‘impossible worlds’. The point which Wittgenstein surely had right is that there is something fundamentally mistaken about the thought that there might be any sort of contrast to be had here: there is no such notion as that of a logically impossible world.

2 Implicationism all the way down?

One of the complaints commonly made by mathematicians about the philosophy of mathematics is that it seems to concentrate so much on three special cases, arithmetic, analysis and set theory. Don’t philosophers realize, they ask despairingly, that most of modern mathematics is not about these but about abstract structures — groups, Hilbert spaces, infinite-dimensional manifolds, Lie algebras?

Well, yes they do realize this. The reason they focus, and this article will focus, on the three special cases is not that they are the only cases philosophers know about (although that may be true of some) but because what makes these cases special is also what makes them specially problematic. If I prove something about all groups, or all Lie algebras, then what I know *is* just a piece of logical knowledge. The philosophical view known as implicationism (or, less elegantly, if-thenism) is a perfectly adequate explanation of what is going on in these cases.

Notice, though, that the theory of groups, or the theory of Lie algebras, is a matter not just of proofs but of counterexamples, and something more needs to be said about them. If we want to show that not all abelian groups are cyclic, what we do is to construct a model of a non-cyclic group, such as the Klein four-group. We cannot (at least without further explanation) treat this in an implicationist manner, since the conclusion we want is not conditional.

In a simple case like the one just mentioned we do not (or at least not consciously) invoke a background theory in which to construct the model: to construct a Klein four-group, for instance, it would be enough to point to the symmetries of a rectangle. In more complicated examples, however, mathematicians do, when constructing counterexamples, invoke the idea that there is a background theory in which the construction is carried out. If they are asked what that background theory is, many of them will say it is **ZFC**. But in most cases it is very doubtful whether they really mean this. Most of them, most pure mathematicians that is, have only an approximate idea what **ZFC** is. And in constructing counterexamples they make no attempt to exploit the power or the intricacies of **ZFC**. The only counter-instances to this that I can think of occur in parts of mathematics that really are quite close to set theory (most notably general topology).

This is an instance of a phenomenon any philosopher of mathematics, at least any who takes the practice of mathematicians at all seriously, has to get used to, namely that what mathematicians *say* is not always a reliable guide to what they are doing: what they mean and what they say they mean are not always the same. The deference towards **ZFC** which many mathematicians claim is really a myth. They do not really believe that the criterion of correctness of a putative theorem is whether it can be formalized in **ZFC** (as is shown, rather trivially, by the fact that they can be persuaded of the truth of Con(**ZFC**)).

Of course, I do not mean by this that there is some *other* formal theory which provides the context for mathematical theorizing of this sort. My point is only that although when constructing counterexamples mathematicians situate themselves in a background context, we should not too readily understand that context to consist in some first-order formal theory.

The central point we need to observe about this background is that we cannot be implicationist about it as well, because that simply postpones the problem: at some point in the process there needs to be something we can assert as *true*, not just conditionally.

In the 1920s it was thought for a while (by Hilbert, most famously) that proof theory might provide a way out. In order to show that something follows from some premises, we prove it. To show that it does not, we might hope to analyse the combinatorial possibilities encapsulated in the rules of proof of the formal system and thereby demonstrate that no string of symbols constitutes a proof in that system of the proposition in question. But this turned out to be a blind alley: in non-trivial cases the proof-theoretic analysis never delivers the required result without making assumptions about transfinite induction which require a substantial background context of their own.

3 Mother theories

So focus from now on on arithmetic, analysis and set theory. There is a key difference between these theories and the theories of groups, Hilbert spaces, etc. that we considered earlier. This difference is that we ordinarily conceive of arithmetic, analysis and set theory as having a unique intended model (the natural numbers, the real numbers, the sets). Of course, we should not pre-judge the philosophical issues simply by assuming that the conception we ordinarily have of these disciplines is correct. Nonetheless, we need to recognize something this conception we have of arithmetic, analysis and set theory entails for our philosophical account, namely that any account of a broadly implicationist shape will have a difficulty with detaching the antecedent of the conditional.

When I talk here of accounts with 'a broadly implicationist shape', I intend to include not just implicationism proper but axiomatic formalism, structuralism and fictionalism. Where they differ is in the form of the antecedent to be detached, not in the need to detach it. The axiomatic formalist says that terms like '7' and '5' gain their meanings from the role they play in the Peano-Dedekind axioms, and those axioms entail that 7+5=12. The modal structuralist says that if there were a natural number structure it would be the case in that structure that 7+5=12. The fictionalist says that according to the story of arithmetic 7+5=12. But in each case they owe us an explanation of how it follows from their account of the matter that seven apples and five oranges make twelve pieces of fruit.

Of course, proponents of these views all do try (with varying degrees of success) to give such an explanation, but they all need to establish the consistency of the axiomatization as a precondition for the success of the explanation. (Typically they need a little more than bare consistency, but we need not go into that at present.) And the difficulty with this is that nothing internal to the theory provides them with the resources to establish this. So they need an external explanation for consistency. But where is that to come from?

Let us call this the *postponement problem* and the positions that face it collectively *postponement* views.

The familiar situation, of course, is the one where we model one mathematical theory in another, but, as we observed earlier, we cannot use that now because it would only postpone the difficulty. What we have to show is that the theory in question is consistent, not merely that it is consistent if some *other* theory is true. When it comes to the mother theory, that is to say, there is no further background in which to do the modelling: it *is* the background.

At this point, incidentally, nothing much hinges on whether the mother theory counts in the traditional taxonomy of these matters as *mathematical*. Thus, for instance, some postponers have justified the consistency of mathematics by appealing to possible worlds in which mathematics is true: for them the theory of modality is the mother theory, and they now owe an exactly analogous debt in respect to it.

The important thing to see is how difficult the postponement problem is for the views that face it. The central difficulty is that the consistency of a substantive mathematical theory is not a trivial matter. This is graphically illustrated by Gödel's incompleteness theorem, which tells us that the consistency of Peano Arithmetic cannot be proved in Peano Arithmetic itself. One suggestive way of thinking of this is that from the perspective of the theory itself the problem of proving its consistency is infinitely hard. Of course, it does not follow that from *every* perspective the problem is hard.

And indeed non-postponement views have a simple argument for consistency: the theory is consistent because it is true in its intended model. But for them the complexity is transferred to the issue of what is involved in grasping the intended model. For non-postponement views, on the other hand, it is intrinsically unlikely that the task of establishing consistency of the theory will be easy, because they are trying to do this from scratch. I have made this point elsewhere in relation to one particular postponement view (neo-Fregean logicism), but it applies equally to all of them.

4 Understanding and truth

I want to make here two distinctions which I think are useful. Both are obvious and very familiar to epistemologists, but I think considering their implications will take us a surprising distance with the question we are dealing with here.

The first distinction is that between understanding and truth, between what is involved in understanding a sentence of mathematics and what is involved in coming to know that it is true.

I promised earlier that the distinctions I was drawing would be obvious ones, but in the history of philosophy this distinction has not always seemed obvious. Verificationism was the attempt to remove it, and regard the meaning of a sentence as consisting simply of conditions under which we might come to know it. Philosophers have, I think, come quite widely to see that this is mistaken, but I do not have to address the issue here in its full generality. It is enough for current purposes to notice that in the case of mathematics the view is especially unpromising. Wittgenstein in his middle period, when he was in the grip of verificationism, did indeed try out the idea that the meaning of an arithmetical generalization consists in its proof, but he never made any real progress with the twin problems that this view faces: on the one hand of explaining how we apparently understand arithmetical sentences, such as Goldbach's conjecture, which we currently have not the least idea how to prove; and on the other of what to say about cases where we have two completely different proofs of the same theorem.

So in the case of arithmetic, for example, there are two distinct issues, one of explaining our grasp of the concepts involved, such as addition, the other of explaining our knowledge of the truths we express using those concepts.

Notice, though, that the two issues are to some extent inter-related, since which concepts we have is partly dependent on what we know. Of course which concepts we have depends to some extent on our environment, both on what is there and on what we find salient. So of course reflection on our concepts is one route by which we can come to some (fairly limited) knowledge about that environment. If there was not, and had never been, any water, we would not have the concept of water. So someone who reflects on his grasp of the concept of water can thereby come to know that there is water. If there were not, and have never been, any antelopes, we would not have the concept of an antelope. So someone who reflects on his possession of the concept of an antelope can thereby come to know that there are such things as antelopes. Someone who, not realizing that unicorns are mythical creatures, uses as similar argument to reach the conclusion that there are such things as unicorns is simply making a mistake.

To those wrapped up in recent the recent debate about the so-called armchair knowledge problem (w Davies 2000, Brown 2001, Beebee 2001, e.g.)hat I have just said will probably seem much too swift. I cannot say much more than that it does not seem too swift to me. The particular cases I have mentioned are of course strictly irrelevant to the issue at hand here, since our knowledge that there is such stuff as water is not mathematical. What is relevant is only to note that there are such things as *armchair concepts*, concepts, that is to say, reflection on our grasp of which is capable of yielding non-trivial knowledge. *Water* is, I claim, one such concept. There are others: Kripke's example of the length of the metre stick in Paris shows that *the metre* is another. What relevance this notion of an armchair concept has to the philosophy of mathematics will emerge later.

5 The route to knowledge

The second distinction I want to draw is that between how I have come to know something and an original or primary route by which it might become known.

If you ask how I came to know some mathematical truth, the answer will almost always be a philosophically disappointing one, namely that I read it in a book, was told it by a teacher, or saw it on the screen of a calculating machine. And I am not unusual in that: those will be the answers almost all of us give for almost all pieces of mathematical knowledge. The route to mathematical knowledge is almost always, in other words, a posteriori.

Nonetheless it is common to insist that mathematical knowledge is a priori. What is meant by this, of course, is not that I in fact came to it independent of experience, or even that someone at some time came to it independent of experience, but only that someone could have done.

Notice, though, how bound up this notion is with issues of modality. Many mathematical theorems that the mathematical community regards as known have proofs which, if written out completely, would be far too long for any one person to have had a full grasp of them. The most that one might claim for such proofs is that they consist of chains of subproofs, each of which has been, at least temporarily, grasped by some people (usually the author of the paper and the referees). So if we describe these theorems as knowable a priori, we probably do not mean that anyone really could come to know them independent of experience. The modality here is not a practical one, but a question of what an idealized reasoner could come to know.

6 Benacerraf's problem

One way of posing the problem of mathematical knowledge that has become standard is due to Paul Benacerraf. So standard has it become that it is nowadays a painful cliché for articles on mathematical epistemology to begin by stating 'Benacerraf's problem' . What we want, it is said, is a naturalistic account of the epistemology of mathematics, and any such account will involve a causal connection between the knower and the objects known about. But the surface syntax of mathematical statements makes it seem that the terms in them refer to mathematical objects. Mathematical objects are abstract and therefore cannot participate in causal chains. So we cannot know about mathematical objects.

Thus Benacerraf's problem. But it seems to me to be a thoroughly misleading way of putting the issue, and to encourage thoroughly unhelpful ways of thinking about it. Let me explain why. Note first that it is actually quite hard to make the view precise in such a way as to explain what role the mention of *objects* is playing. The reason for this is that aboutness is much more slippery than people tend to think. E.g. what is the law of supply and demand in economics about? Physicalist will have to say it is really about electrons and protons. But which electrons and protons? What about future configurations, or possible configurations? Consider 'Either it's raining or it isn't'. What is this about? According to Wittgenstein's theory in the *Tractatus* it is not about anything. Is that right?

Recalling earlier distinction between the meaning of a sentence and what is involved in verifying it, note that there will correspondingly be two notions of aboutness, i.e. first what is involved in grasping the proposition and second what is involved in coming to know it. It is presumably the second of these that is relevant to Benacerraf's problem, and for this it is not obvious that there need be anything a proposition is about: different routes to verifying it may involve different objects, and there may be no intersection between them.

I am not denying that there is *a* sense in which 7+5=12 is about the number seven (among other things). What I am denying is that this sense need be the only one relevant to epistemology. (Compare the sentence about the rain. To know that it is true I don't need to have ever seen rain. What I need to know is only that rain is a concept in good standing of a particular sort. But that might be regarded as purely grammatical information.)

Suppose, though, that we could resolve these issues satisfactorily, and it emerged that there is a moderately stable sense of aboutness according to which we can identify at least some of the objects mathematical sentences are about. Even then it would not clear what role the objects should be expected to play in causation. To make an obvious point, the relata of causation may be facts or events, but what they are not is objects: talk of objects as either causes or effects is at best a loose way of talking and at worst a category mistake.

So to ask how abstract objects can be causes is misleading: what is meant is the facts or events that the abstract objects are involved in. But if you believe in abstract entities at all, you surely cannot think that the facts or events that are causes should be expected to have *no* abstract constituents to them. So the problematic case can only be that of facts or events *all* of whose constituents are abstract, i.e. ones with no concrete constituents at all. But remember that what we are after here is a problem for *mathematical* knowledge that is not also a problem for logical knowledge. So the problematic case has to be that of a cause all of whose constituents are abstract and at least one of which is an object. But there is work to do to explain why case this should be thought especially problematic.

Suppose, though, that we grant all this. Why think that there has to be any causal connection between a fact involving an object and the event that constitutes my coming to know this fact (or the fact that constitutes my knowing it). The *locus classicus* for the claim that there has to be such a connection is W. D. Hart:

> Granted just conservation of energy, then, whatever your views on the mind-body problem, you must not deny that when you learn something about an object, there is a change in you. Granted conservation of energy, such a change can be accounted for only by some sort of transmission of energy from, ultimately, your environment to, at least proximately, your brain. And I do not see how what your learned about that object can be *about* that object (rather than some other) unless at least part of the energy that changed your state came from that object. It is all very well to point out that the best and (thus) true explanation of our state changes in learning probably requires the postulation of objects, like numbers, which cannot emit energy, but about which we nevertheless have beliefs. For this still leaves unexplained how our beliefs could be about energetically inert objects. (Hart 1977, p. 125)

But as it stands, this is hopeless. To take just one obvious example, there are quite a few things I know about various objects outside my light cone. I know, for instance, that they (or many of them) obey the laws of physics (at least approximately). Consider tomorrow's sunset. There is no causal chain from this event to my current state, and yet I know various mundane things about it — when it will happen, in which direction from where I am now, perhaps even approximately what it will look like.

The standard response to this, of course, is that we should not demand, as Hart does, a chain *from* the event known about to the current state of the knower, since that rules out knowledge of particular events in the future. It may suffice, at least in appropriate circumstances, for the event known about and the state of knowledge to have a *common* cause. But it is very hard indeed to see, in the mundane cases we are considering now, what the common cause is. Much of my knowledge about future events, such as tomorrow's sunset, is knowledge that I obtain by application in particular instances of general laws. I know (or can look up in my diary) a general law for predicting the times of sunsets, and I apply it in the particular case at hand. But the law written down in my diary is not the cause of tomorrow's sunset.

7 Benacerraf's problem generalized

So much the worse, one might say, for a causal theory of knowledge. But without such a theory, what is left of the thought that there is anything especially problematic about knowledge of abstract objects? At this point the inheritors of the Benacerrafian tradition are apt to generalize the difficulty. It is not, they say, a problem about causal knowledge but merely about natural knowledge.

> It is a crime against the intellect to try to mask the problem of naturalizing the epistemology of mathematics with philosophical razzle-dazzle. Superficial worries about the intellectual hygiene of causal theories of knowledge are irrelevant to and misleading from this problem, for the problem is not so much about causality as about the very possibility of natural knowledge about abstract objects. (Hart 1977, pp. 125-6)

Now 'naturalism' is an often-used word in recent philosophy, but what it means is sometimes not as clear as it should be. One thing that is striking about the way it is used, for instance, is that quite often it really seems to mean physicalism. But I think we can put that behind us quite quickly. If *that* is what is meant, why on earth should we believe it? Why, that is to say, imagine that the methods of physics provide any sort of guidance as to the sort of epistemology we should adopt?

So let us suppose from now on that it is a broader naturalism we mean. Naturalism has sometimes been offered as a guide in issues connected with ontology: all there is is what scientists tell us there is. This is usually interpreted to mean 'scientists speaking in their role as scientists'. But that seems to mean 'scientists speaking within their sphere of expertise', which begs the question of what their sphere of expertise is. Physicists are the experts on what physical objects there are. Biologists are the experts on what creatures there are. Are theologians the experts on what Gods there are?

Whether we extend the argument to theologians depends on whether they are scientists. On the face of it this is absurd. But why? One answer sometimes given is that there are distinctively scientific norms which e.g. theologians (or literary critics) do not adhere to, but which natural scientists (when performing their roles as natural scientists) do adhere to. What are they?

Experiment certainly has a role in some sciences. But experiment entails interaction. 'If you can spray them then they are real.' (Hacking 1983, p. 23) And that excludes quite a few of what are usually categorized as sciences. Economics is surely a good example. One could, of course, take the heroic course of denying that economics is a science, but that does seem ill-advised: it would be foolish to deny, for instance, that there is a great deal of knowledge encoded in economic theories. And what, then, of astronomy? (Apart from anything else, this shows that Hacking's slogan is far from a criterion for reality: we do not think that our inability to manipulate distant stars shows that they are not real.)

So if experiment is not characteristic of the natural sciences, what about observation? This certainly plays a role even in astronomy and in economics. But isn't it now too broad a categorization? Surely observations (or at any rate empirical data) are relevant not just in science but in any activity whatever.

The general point is that it is quite difficult, I think, to characterize the natural sciences by finding distinctively *scientific* norms which they all adhere to. When one tries to formulate them, what one comes up with tend merely to be *rational* norms. Not everything that is rational is part of the natural sciences, and what distinguishes the sciences from other forms of rational enquiry is more to do with the subject matter and the company scientists keep than with anything distinctive about the *epistemological* norms they adhere to.

The contrary view seems to me to have been borne of a false contrast according to which what is opposed to science is 'mere' — mere astrology perhaps, or mere theology. We may agree that astrology is in this sense mere: the claims made in astrology are not knowledge. But it is much too big a jump to say that everything that is not science is in astrology's boat.

8 Generalized again?

One might be tempted to generalize Benacerraf's problem still further (and, thereby, give it a much older heritage, reaching back to Plato's *Meno*). Perhaps, one might say, the problem is not about whether the epistemology of abstract objects deserves to be called naturalistic but only about whether there is any sense in which it tracks the objects at all. No belief we possess deserves to be called knowledge, we might say, unless it covaries counterfactually with what is known: if it were not true, I would not know it. And the problem for mathematical knowledge is just that it does not vary: the counterfactual gets no grip because if a mathematical sentence is true we cannot make the required sense of supposing it not to be.

But that will not do for current purposes, simply because the problem has now been generalized to the point that it applies to *any* necessary truth. It no longer holds any terror for the mathematician that it does not also hold for the logician.

9 The real problem

What, then, is the real problem of mathematical knowledge?

It will be clear from what I said earlier that I think the postponement problem is a serious one for the views which face it (what I called postponement views); indeed, I think that it is fatal, although I have not said enough here to demonstrate that. The reason that non-postponement views do not face this problem is of course because they have a quick answer to the question why the theories they are dealing with are consistent: they are consistent because they are true about the intended model.

The real problem for non-postponement views is therefore to explain what a conception of a mathematical structure amounts to in such a way as to make it plausible that we might come to know that certain things are true about it. But, as my scepticism concerning Benacerraf's problem should make clear, I do not think that what makes this especially problematic is that we conceive of the structure as abstract.

That is not to say, of course, that knowing about numbers is just like knowing about tables and chairs, or even that it is just like knowing that either it is raining or it isn't. But the point at which the epistemological problem becomes one that is distinctively mathematical comes when we invoke the idea of reflection.

This is why I mentioned armchair knowledge earlier. What the examples we saw there showed was that there are concepts, armchair concepts I called them, our possession of which entails facts which are neither trivial nor in any ordinary sense about us. Our possession of the concept *water*, for instance, entails that there is water. What I want to claim is that some mathematical concepts, most prominently arithmetical and geometrical concepts, are armchair concepts in this sense.

One might be tempted to say not just 'in this sense' but 'in this way'. That would be too strong, however. That *water* is an armchair concept follows from semantic externalism. And semantic externalism about this concept is forced by consideration of how our concepts would differ if we lived (and had always lived) on Twin Earth. That *number* is an armchair concept does not follow from semantic externalism: one cannot be a semantic externalist about the concept *number*, because the counterfactual account of what semantic externalism amounts to gets no grip on it. Any attempt to explain platonism about mathematics by means of a counterfactual anything like the Twin Earth case is lame from the start.

The challenge which mathematics presents to the epistemologist is therefore to explain how *number* is an armchair concept.

References

Beebee, H. (2001). Transfer of warrant, begging the question, and semantic externalism. *Philosophical Quarterly 51*, 356–74.

Brown, J. (2001). The reductio argument and transmission of warrant. In S. Nuccetelli (Ed.), *The reductio argument and transmission of warrant*, pp. 117–30. Cambridge, MA: MIT Press.

Davies, M. (2000). Externalism and armchair knowledge. In P. Boghossian and C. Peacocke (Eds.), *New Essays on the A Priori*. Oxford University Press.

Hacking, I. (1983). *Representing and Intervening.* Cambridge University Press.

Hart, W. D. (1977). Review of steiner, mathematical knowledge. *Journal of Philosophy 74*, 118–29.