3

Is mathematical knowledge just logical knowledge?*

Logicism is the claim that mathematics is part of logic. This claim flies in the face of Kant's denial that mathematics is analytic, that is, true by logic and definitions alone; and it seems to me that if mathematics is taken at face value, Kant is surely right.

The reason for this assessment is that mathematics, if taken at face value, makes existential assertions: it asserts for instance that there exist prime numbers greater than a million, and therefore that there exist numbers. Indeed, even Kant's example '$5 + 7 = 12$' makes an existential assertion, if understood in the usual way: it asserts not only that *if* there are x, y and z such that $x = 5$ and $y = 7$ and $z = 12$ *then* $x + y = z$, but the further existential claim that there *are* such x, y and z. So to argue against the idea that mathematics, if taken at face value, is true by logic and definitions alone, we only need argue that you can't get existential assertions out of logic and definitions alone.

And Kant did provide such an argument (though not in his discussion of mathematics). Anselm, Descartes and others had argued that the

* I would like to thank Tony Anderson, Paul Benacerraf, Charles Chihara, Geoff Joseph, Janet Levin, David Lewis, Penelope Maddy, Colin McGinn, Bill Tait, Richard Warner and an anonymous referee, for helpful comments that have affected the final version. Much of the research was carried out under a summer research grant from the National Science Foundation (SES-8205264).

Note to this volume: I have done some rewriting of this paper. Apart from section 4, the only substantive revision is in the last paragraph of fn. 15. Section 4 has been considerably rewritten, in an effort to clarify my position and remove some minor errors. The main change is that in the original version of the paper I introduced a 'broad conception of standard mathematics', replete with substitutional quantifiers and an ω-rule for them; but I have come to see that my point could be made more clearly without this. There are also changes in my remarks on nominalistic proof theory. Not everything that I would change if I were writing the paper today has been changed: see the postscript for additional remarks.

existence of God is a matter of logic, of conceptual necessity: that it follows from the very concept of God that God exists. Kant argued that this can't possibly be correct, for logic (and logic together with definitions) can never categorically assert the existence of anything. Kant's argument for this principle is that contradictions usually stem from postulating one or more objects and making various assumptions about the postulated object or objects that are mutually inconsistent: for example, postulating a triangle, and then saying something else that implies that it has more than three sides. But there is never a contradiciton if we reject the triangle – there is nothing there about which we have made contradictory assumptions. And, to quote Kant, 'the same holds true of the concept of an absolutely necessary being. If its existence is rejected, we reject the thing itself with all its predicates; and no question of a contradiction can then arise.'[1] He sums it up by saying 'I can not form the least concept of a thing which, should it be rejected with all its predicates, leaves behind a contradiction.'[2] I think that this argument is rather persuasive. If it is correct, it cannot be contradictory to deny the existence of God; and it cannot be contradictory to deny the existence of numbers either, for they don't have the mysterious power of leaving behind a contradiction when their existence is rejected any more than God does.

One can quibble with this argument of Kant's for the principle that logic and definitions alone imply no existence assertions; nevertheless, the principle itself is a very compelling one. Perhaps when a person denies the existence of God or of numbers, what the person is saying is false or even 'metaphysically impossible' (whatever that means); but it is not itself a logical contradiction in any normal sense of 'logical contradiction'. Moreover, there is good reason not to depart from the normal sense of 'logic' by counting existence assertions as part of logic: doing so would tend to mask the fact that there is a substantive epistemological question as to how it is possible to have knowledge of the entities in question (God, numbers, etc.).[3]

1 *Critique of Pure Reason*, B622–3.

2 *Ibid.*, B623–4.

3 Admittedly, there are also questions about how it is possible to know logical truths like 'If snow exists then snow is white or snow is not white'; but such questions seem much less gripping than questions about how we can know the existence of specific kinds of entities, and it seems very unlikely that any reasonable answer to the question of how logic is known would bring with it an answer to the question of how the existence of God or of numbers or of any other specific sorts of entities is known. See for instance Paul Benacerraf, 'Mathematical truth' for a discussion of epistemological difficulties that our apparent knowledge of the existence of mathematical entities raises and which don't seem to be raised by straightforwardly logical knowledge. (In essay 2 I discuss this further and argue that contrary to what is sometimes claimed, the epistemological difficulties that

So mathematics, taken at face value, can not be reduced to anything reasonably called logic. In a sense, of course, the classical logicists did take mathematics at other than face value: they held that though it at first blush appears to concern specifically mathematical entities like natural numbers, real numbers and tensors, it can really all be shown to be part of the theory of properties (or the theory of propositional functions, or the theory of extensions of concepts, or whatever). But the theory of properties, or propositional functions, or whatever to which the classical logicists hoped to reduce mathematics asserted the existence of a vast array of properties (or propositional functions), so the problem of how this can reasonably be regarded as logic recurs. (It is, indeed, a problem which eventually led one famous logicist, Frege, to abandon his logicism: 'it seems that [logic] alone cannot yield us any objects . . . [So] probably on its own the logical source of knowledge cannot yield numbers'.[4] If one is going to retain logicism (in conjunction with the Kantian requirement on logic that I have advocated), one is going to have to provide an interpretation of mathematics according to which mathematics does not really make existential assertions despite all appearances to the contrary. I do not regard the prospects of doing this in a plausible way as at all promising, so I will not be defending logicism.

Still, I think that the idea that mathematical knowledge is just logical knowledge is largely correct, for I want to maintain what might be called a *deflationist* position about mathematical knowledge. That is, I want to say that what separates a person who knows a lot of mathematics from a person who knows only a little mathematics is *not* that the former knows many and the latter knows few of such claims as those that mathematicians commonly provide proofs of (i.e., of those claims, such as that there are prime numbers greater than a million, which I have claimed to be non-logical). Rather, insofar as what separates them is knowledge at all,[5] it is knowledge of various different sorts. Some of the knowledge that separates them is empirical knowledge (e.g., about

---

Benacerraf discusses do not depend on assuming a causal theory of knowledge.) It would hardly be a solution to the problem that Benacerraf raises to say that we know that numbers exist because *logic guarantees* that they exist.

[4] 'A new attempt at a foundation of arithmetic', pp. 278–9.

[5] Another thing that separates those who know lots of mathematics from those who know only a little isn't knowledge at all strictly speaking, it is *ability* ('knowledge-how' as opposed to 'knowledge-that'): the ability to prove mathematical theorems, the ability to see the relevance of mathematical theorems to practical matters, and so forth. But only to the extent that the possession of such abilities depends on the possession of knowledge-that does the possession of such abilities raise epistemological problems; that is why in the text I have focused only on the *knowledge* (knowledge-that) which separates those who know lots of mathematics from those who know little.

what other mathematicians accept and what they use as axioms). Putting empirical knowledge aside, my claim is that the rest of the knowledge that separates those who know lots of mathematics from those who know only a little is knowledge of a purely logical sort – even on the Kantian criterion of logic according to which logic can make no existential commitments.

The epistemological advantages of such a view are obvious: it obviates the need of postulating mathematical knowledge that is not logical and hence that is presumably synthetic *a priori*; and (putting questions of *a prioricity* or *a posterioricity* aside) it obviates the need for postulating epistemological access to a special realm of mathematical entities. Nonetheless, it is not at all obvious how any such deflationist view is to be worked out in detail, or how plausible it can be made. This paper is an attempt to survey some of the main problems that must be overcome in defending a deflationist view and to suggest ways of dealing with them.

1

The crudest attempt to state a deflationist position would be to say that all mathematical knowledge is really just knowledge that certain mathematical claims follow from certain other mathematical claims and bodies of claims: we can know, for instance, that the claim that there are primes greater than a million follows from the usual axioms of number theory. (This form of deflationism is reminiscent of, but importantly different from, what is sometimes called 'deductivism' or 'if-thenism'. Some of the differences will be discussed near the end of the essay.) This crude form of deflationism is difficult to believe: for in addition to knowing that certain claims follow from certain bodies of other mathematical claims, don't we also know the consistency of some of those bodies of mathematical claims? For instance, don't we know the consistency of various mathematical axiom systems? If one were to take the crude form of deflationism seriously, one would have to say that we can't really know that an axiom system is consistent: we can know only that the consistency of one axiom system follows from the axioms of another system which itself can't be known to be consistent (though of course its consistency can be known to follow from still a third axiom system). This strikes me as extremely implausible.

Fortunately, there is no need for a deflationist to be a crude deflationist, and so no need for him or her to take this line on consistency. For it would seem that knowledge that a certain axiom system is consistent (i.e., knowledge that some claim of the form 'p & -p' *doesn't follow* from the system) is every bit as much *logical* knowledge as is knowledge that a certain claim *does follow* from the system. The implausibility of the crude form of deflationism lies in its being forced to try to explain apparent knowledge of *what doesn't follow* in terms of knowledge of *what does follow*. But there is no point in trying to do this if both have equal claims to count as logical knowledge.[6]

A less crude form of deflationism, then, is the view that the only knowledge that differentiates a person who knows lots of mathematics from a person who knows only a little (aside from empirical knowledge of various sorts, such as the sort mentioned earlier) is

(i) knowledge that certain mathematical claims follow from certain other mathematical claims or bodies of claims,
(ii) knowledge of the consistency of certain mathematical claims or bodies of claims

and other knowledge of a basically similar sort; and that all this knowledge is logical.

Unfortunately, however, there is a powerful objection to this less crude form of deflationism (and to the cruder form as well); and that is that the knowledge cited in (i) and (ii) of the previous paragraph is *not* logical knowledge. For instance, the knowledge in (i) – that certain mathematical claims follow from certain others – isn't *logical* knowledge (i.e., knowledge of logical truths); it is *metalogical* knowledge, for it is knowledge *about the relation of logical consequence*. Now, the relation of logical consequence can be understood in either of two ways. It can be understood semantically; in that case, to say that A is a logical consequence of T is to say that in all models in which T is true, A is true as well. But so understood, the knowledge that A is a consequence of T is knowledge about *all models*. Models are mathematical entities, so knowledge about logical consequence understood semantically is a special sort of mathematical knowledge. It is not knowledge of a logical truth, such as 'If snow exists, then snow is white or snow is not white.' So much for the semantic construal of logical consequence; but how about the syntactic construal, on which to say that A is a logical consequence of T is to say that there is a formal derivation of A from T (according to the rules of some specific formal system)? Clearly this is no better, for then knowledge that A is a consequence of T in the syntactic sense is knowledge *of the existence of formal derivations*. This cannot be logical knowledge: logic can't assert the existence of formal

[6] As we will see in the next section, there are grounds which could lead some people (though they won't lead me) to deny that both have equal claims to count as logical knowledge.

derivations any more than it can assert the existence of God. Indeed, it is mathematical knowledge, for formal derivations in the intended sense are the abstract objects dealt with in the mathematical theory of proof. (They aren't simply strings of symbols on paper, for A can be a consequence of T without there being a piece of paper on which someone has written a formal derivation of A from T.)

I have stated the objection as an objection to the claim that the kind of knowledge mentioned in (i) is logical knowledge; but clearly the objection holds equally well for the kind of knowledge mentioned in (ii). One might be tempted to conclude that the idea that there is no mathematical knowledge over and above logical knowledge is simply a mistake.

I want to resist this conclusion, and to see how to do this, it is worth noting that even independently of the fact that you seem to need mathematical entities to define logical consequence and consistency, there is something else unintuitive about the idea that the only mathematical knowledge there is is strictly speaking of form (i) or (ii) or something similar. For the knowledge cited in (i) and (ii) is metalinguistic; are we really to hold that all mathematical knowledge is metalinguistic? Indeed, doesn't the metalinguistic fact that one sentence follows from another depend on the fact that certain words appearing in these sentences are used as logical words by speakers of the language in question? If so, knowledge of what follows from what has a contingent element that the mathematical knowledge we were trying to convey presumably lacks. This then is another (admittedly less compelling) objection to the version of deflationism under discussion. And it seems clear that we can get around both objections simultaneously if we can find a way to 'semantically descend', that is, to state the sort of mathematical knowledge that the deflationist wants to focus on without going metalinguistic.

How this is to be done is clearest in the case of finite bodies of mathematical claims, and for the moment I will confine my attention to them. If A is the conjunction of all members of a body T of mathematical claims (e.g., the conjunction of all the axioms of a finitely axiomatized theory), then instead of saying that we have mathematical knowledge that this theory is consistent, why not simply say

(ii′) we know that $\Diamond$A

(where the modal operator '$\Diamond$' is to be read 'it is logically possible that' or 'it is logically consistent that'). Here the claims in T are used, not mentioned, so the contingency objection doesn't apply. And because they are used, not mentioned, the symbol '$\Diamond$' cannot be understood as a predicate that needs explanation in set-theoretic or proof-theoretic terms; it must be understood as an operator, and indeed an operator that is widely regarded as an operator of logic; consequently the earlier objection doesn't apply either.

The point I have made for (ii) applies of course to (i) as well. That is, instead of saying that the claim B follows from the body of claims whose conjunction is A, why not say

(i′) we know that □(A ⊃ B),

where '□' ('it is logically true that') is of course defined as '-◇-'.

So our third version of deflationism (the final version, apart from a slight alteration designed to handle theories that are not finitely axiomatized) is that what differentiates a person with lots of mathematical knowledge from a person with only a little (apart from differences in abilities (cf. footnote 5) and in empirical knowledge) is that the former but not the latter has lots of knowledge of the type (i′) and (ii′) and other knowledge of a basically similar sort. (One of the things that this last 'hedge' clause covers is modal knowledge not either of the form '◇A' or '□A'; for instance, the conditional knowledge that *if* it is consistent that A *then* it is consistent that B seems like knowledge that a deflationist could perfectly well allow.)[7]

In moving from (i) and (ii) to (i′) and (ii′), I of course have to accept the idea that the notion of logical possibility is an acceptable notion,

[7] It should be noted that the modal knowledge which deflationism allows is knowledge of purely logical possibility – deflationism does not allow knowledge of mathematical possibility in any interesting sense. This makes deflationism very different from the viewpoint that Hilary Putnam calls 'mathematics as modal logic' in 'What is mathematical truth?'. According to 'mathematics as modal logic',

> the mathematician . . . makes no existential assertions at all. What he asserts is that certain things are possible and certain things are impossible – *in a strong and uniquely mathematical sense of 'possible' and 'impossible'*. (p. 70, my italics.)

Putnam claims that despite its making no existential assertions, mathematics as modal logic really states the same facts as does mathematics on its platonistic interpretation. I think that there is considerable plausibility in Putnam's claim. The reason is that in the 'strong and uniquely mathematical sense of "possible"', 'it is possible that A' is an object-level analogue of 'A is consistent with any true mathematical theory'; something very akin to mathematical truth (and therefore to mathematical existence) is being sneaked into Putnam's possibility operator. Putnam's position apparently is that if you take ordinary set theory S and some nonstandard alternative S′ that is inconsistent with S (but internally consistent), then S is mathematically possible but S′ isn't. The deflationist viewpoint is different: S and S′ are on par in that $\Diamond Ax_S$ and $\Diamond Ax_{S'}$. Of course, S may be more useful than S′ for various purposes, but if so, this requires an explanation. (See sections 3 and 4 of this essay.) A deflationist cannot regard it as an acceptable explanation to say that S describes a mathematical possibility and S′ does not. I believe that Putnam's view is that such an explanation *would* be acceptable and, indeed, would be the only explanation possible.

and also I must accept that it really is a part of logic and not something that must be explained in terms of *entities* (e.g., models, formal derivations or possible worlds). In addition, in accepting that (ii′) counts as logical knowledge (for suitable assertions A), I have to interpret the idea of logical knowledge (and logical truth) in a way broader than current orthodoxy permits. This I will now explain.

2

Consider the claim

(1) $\Diamond \exists x \exists y\ (x \neq y)$,

which says that it is logically possible that there be at least two objects in the universe. Is this a truth of logic? I think the natural answer is 'yes', and most of the people I have asked agree. Nonetheless, on the usual approach to defining logical truth for modal logic (Kripke's approach),[8] (1) does not come out logically true. Indeed, it is a curious feature of Kripke's approach to defining logical truth for modal logic that *no* sentence of the form '$\Diamond$A' is logically true, except in the trivial case where A itself (and hence $\Box$A) is logically true. To me, this seems quite unmotivated. It may be countered that while it is perhaps initially natural to regard (1) as a logical truth, Kripke's model-theoretic definition of logical truth for modal logic is also quite intuitive, and it is a consequence of this model theory that (1) not be logically true. My reply is that there is an alternative model-theoretic definition of logical truth for modal logic that is simpler and I think more natural than Kripke's, and which is much closer than Kripke's to the model-theoretic definition of logical truth for first order logic (for instance, in involving no reference to 'possible worlds' or to any other entities not used in the model theory for first order logic); and according to this alternative model-theoretic definition of logical truth, (1) comes out logically true. (The basic idea of this alternative method of defining logical truth has occurred to quite a few people, starting with Carnap. I describe the version of it that I favour in an appendix).

Indeed, it is a consequence of the non-Kripkean approach to defining logical truth for modal logic that any assertion of the form '$\Diamond$A' that is true is logically true, and that any assertion of the form '$\Diamond$A' that is false is logically false. It is essential to the plausibility of this that '$\Diamond$' be read 'it is *logically* possible that' – it is *logical* possibility (not mathematical possibility or 'metaphysical possibility') that we are concerned to give the logic of. (See footnote 7.) Exactly what the force

8 'Semantical considerations on modal logic'.

of 'logically possible' is depends on further stipulation: it depends on what one takes non-modal logic to include. Some philosophers (e.g., Carnap) regard the non-modal logic to which we are adding '$\Diamond$' as including not only first order logic properly so called, but also 'meaning postulates' specifying 'logical' relations among predicates. Consequently, a non-modal sentence such as

(2) $\exists x$ (x is a bachelor & x is married)

would count as *logically* false for Carnap, and as a result,

(3) $\Diamond\exists x$ (x is a bachelor & x is married)

also comes out logically false. But I prefer not to follow Carnap in taking meaning relations among predicates to be part of logic. My preference is not based on any firm doctrines about analyticity; indeed, my preference is partly based on a desire to remain neutral about such issues. Mostly, however, my preference is based on simplicity: it is simpler to develop a basic modal logic that takes no account of meaning relations among predicates; once one has such a logic, it is easy to obtain from it a derivative logic that takes into account any relations among predicates that one cares to regard as meaning-postulates, if one so desires. If one adopts this strategy – and I shall – then (2) is not *logically* false; it is *logically* consistent that there be married bachelors (even though it may not be consistent with meaning-postulates that there be married bachelors). Now, if it is *logically* consistent that there be married bachelors, and '$\Diamond$' is read as an operator meaning 'it is *logically* consistent that', then (3) comes out true. Indeed, any sentence of the form '$\Diamond\exists x$ (Px & Qx)', where P and Q are atomic, comes out true; what else could you expect if meaning relations among predicates are not taken into account? Moreover, it seems as if it ought to be part of the *logic* of the logical consistency operator '$\Diamond$' that sentences of the form '$\Diamond\exists x$ (Px & Qx)' are true. That is,

(4) $\Diamond\exists x$ (x is red & x is round)

should be not only true but logically true; and similarly for (3). The view that (3) shouldn't be logically true, indeed shouldn't even be true, results from giving to '$\Diamond$' a sense not intended.[9]

[9] As remarked above, there need be no doctrinal difference between the view I have advanced, according to which (3) counts as logically true, and Carnap's view, according to which a sentence typographically like (3) counts as logically false. For we can represent the Carnapian view within the view I have advanced by introducing Carnapian notions by definition into both the object language and the metalanguage. Thus 'is C-logically true' and 'is C-logically false' are to mean 'follows from . . . ' and 'is inconsistent with . . .', where the blanks are to be filled by the 'meaning postulates' for English; and

cont. over page

I hope this gives some idea of (and some motivation for) the view of modality and of the logic of modality that I will be presupposing. For a bit more detail, see the appendix.

Let us return to the issue of mathematical knowledge, and in particular to the version of deflationism arrived at in section 1. Part of the position arrived at there was (a) that mathematicians sometimes know things of form $\Diamond$AX, where AX is a conjunction of axioms of a theory; and (b) that this knowledge is logical knowledge. Now a minimum condition for (b) to hold is that $\Diamond$AX must count as a logical truth. AX itself won't be logically true, if it is a conjunction of axioms of a typical mathematical theory; so in order to adhere to the version of deflationism put forth in section 1 we clearly have to adhere to a non-Kripkean conception of logical truth according to which some non-trivial assertions of possibility are part of logic. The non-Kripkean conception of logical truth sketched above (or if you prefer, the more fully Carnapian variant sketched in footnote 9) will do. Indeed, they have the feature that the modal assertion $\Diamond$AX will be logically true if it is true at all. So there is no danger on these conceptions that we might know that $\Diamond$AX and what we know not be logically true: if we know it, it's true, and so it's logically true.

Does this show that our knowledge that $\Diamond$AX is logical knowledge? Not by itself – for it might be claimed that though $\Diamond$AX is logically true, it is known by non-logical means.

There is a more interesting and a less interesting version of this claim. The less interesting version points out that much of our knowledge of possibility is to some extent inductive. For instance, our knowledge that $\Diamond AX_{NBG}$ (where NBG is von Neumann-Bernays-Gödel set theory and AX_{NBG} is the conjunction of all its axioms) seems to be based in part on the fact that we have been unable to find any inconsistency in NBG. And, it can be claimed, this inductive element in our knowledge precludes that knowledge from being logical. Now, even this less interesting version of the claim that our knowledge of the form $\Diamond$AX is non-logical raises some interesting issues about the nature of logic,

'$\Diamond_C$A' is to mean '$\Diamond$(A & . . .).' Then though (3) is still logically true (and hence C-logically true), still

(3_C) $\Diamond_C \exists x$ (x is a bachelor & x is married)

is C-logically false (and, indeed, logically false). Moreover, we may if we like agree with Carnap that it is C-logical truth and '$\Diamond_C$' that are the philosophically more important notions. Whether or not one agrees with that philosophical claim, the procedure of focusing first on logical truth and on '$\Diamond$' is of quite considerable technical convenience. This will be evident, for instance, when we come to formulate the Conditional Possibility Principle later in this section.

issues that may be relevant to the precise wording of the deflationist's claim. But there is no need to go into such matters, for it seems quite clear that the basic idea of deflationism cannot be undercut by pointing out that much of our knowledge of possibility has a partly inductive character. The basic idea of deflationism is that one is to avoid postulating knowledge of a realm of mathematical entities, and that one is to do this by saying that ordinary mathematical claims are not known to be true. The deflationist holds that what separates those who know a lot of mathematics from those who know only a little is various sorts of knowledge and abilities, none of which give rise to the philosophical problems that knowledge of a realm of mathematical entities gives rise to (or is commonly thought to give rise to). A good deal of the knowledge that separates those with lots of mathematical knowledge from those with only a little is empirical. (I mentioned this earlier, and will discuss it more fully near the end of the essay.) Other of this knowledge is, let us suppose, *straightforwardly* logical in that it involves no inductive elements. And other of this knowledge is knowledge of logical truths by partly inductive means. Perhaps the fact that this latter knowledge is partly inductive keeps it from being logical, and perhaps not. Perhaps it makes the knowledge empirical, perhaps not. I would incline toward answering both of these questions in the negative; but however they are answered, the fact that some of our knowledge of logical truths is partly inductive does not in any way support the claim that it is based on knowledge of a special realm of abstract entities or on knowledge of the truth of ordinary mathematics. Because of this, the fact that some of our knowledge of logical truths is in part inductive can't be used to argue against the essentials of the deflationist position.

As I've remarked, there is also a more interesting version of the claim that though $\Diamond AX$ is logically true, it is known by non-logical means – a version which, if true, would genuinely count against deflationism. Consider what Frege said about knowledge of the consistency of mathematical theories:

> Strictly, of course, we can only establish that a concept is free from contradiction by first producing something that falls under it. (p. 106e)[10]

Obviously this is not literally correct – we can establish that the concept 'horse with wings' is free from contradiction without producing a horse with wings – but the position can be weakened without totally altering its spirit. The weakened version of Frege's claim grants that there is knowledge of possibility that does not arise from knowledge of actuality, but which arises instead from reflection on the logical form of concepts.

[10] §95 of *The Foundations of Arithmetic.*

But, it maintains, all such knowledge of possibility is conditional: one cannot attain categorical knowledge of possibility by this means alone. Rather, categorical knowledge of possibility can only be obtained either directly from knowledge of actuality, or indirectly, that is, from direct categorical knowledge of possibility in conjunction with conditional knowledge of possibility and other logical knowledge. So, for instance, reflection on the common logical form of 'horses with wings' and 'animals with tails' yields the conditional knowledge that *if* it is logically possible that there are animals with tails, *then* it is logically possible that there are horses with wings.[11] This knowledge together with the knowledge that there actually are animals with tails then yields knowledge that it is logically possible that there be horses with wings. The Fregean position is that all knowledge of possibility arises by some such means. (Of course, the knowledge of actuality on which knowledge of possibility is ultimately based may, on Frege's view, be *a priori*.)

If this Fregean position about knowledge of possibility were correct, then deflationism would be in deep trouble. For presumably we know (or at least have good reason to believe) the claim $\Diamond AX_{NBG}$ and the claim $\Diamond AX_R$ where R is the theory of real numbers; but according to deflationism, we do not know (or have good reason to believe) the claims AX_{NBG} or AX_R themselves since they assert the existence of mathematical entities. If Frege were right, then our knowledge that $\Diamond AX_{NBG}$ and $\Diamond AX_R$ would have to be based on conditional knowledge of possibility that arises by reflection on logical form, together with other logical knowledge plus knowledge of actuality. Now, one principle that I think even a Fregean would grant is that if ϕ is non-modal and ψ is a generalized substitution instance of it (i.e., is obtained from ϕ either by substituting formulas for non-logical predicates or by uniformly restricting the ranges of all quantifiers and free variables or both,[12] then we can know that *if* $\Diamond\psi$ *then* $\Diamond\phi$ by reflection on logical form alone.[13]

[11] Here and throughout the rest of this section, the fact that we have not included meaning relations among predicates as part of logic pays off.

[12] The restriction of quantifiers and free variables is to be by a formula $D(x_1)$ which may contain other free variables besides x_1; the formula $A_p(x_1, \ldots x_k)$ to be substituted for the k-place predicate p may likewise contain other free variables. The restriction on quantifiers and free variables is to be made only on the quantifiers and free variables of the original formula, not on any new ones introduced in an A_p or in D. (To be more formal: before performing the general substitution in a formula B, replace all bound variables in B or in D or in an A_p that occur (free or bound) in any other of the formulas by new variables. Then for any sub-formula X of B, associate an X^* as follows: if X is $p(v_1, \ldots, v_k)$, let X^* be $A_p(v_1, \ldots, v_k)$; if X is $-Y$, or $Y \supset Z$, let X^* be $-Y^*$, or $Y^* \supset Z^*$; if X is $\forall v Y(v)$, let X^* be $\forall v(D(v) \supset Y^*(v))$. Finally, the generalized substitution instance $\hat{B}$ of B is $D(y_1)$ & . . . & $D(y_n)$ & B^*, where $y_1, \ldots y_n$ are the variables free *in B*.) The possibility of $\hat{B}$ is in effect the possibility of its existential quantification, both with respect to the variables free in B (now restricted by D) and with respect to any other

If this Conditional Possibility Principle is granted to the Fregean, then by embedding real number theory R in set theory NBG, the Fregean can admit that we can know that

(i) If $\Diamond AX_{NBG}$ then $\Diamond AX_{R}$.

(For if R′ is the set-theoretic assertion to which the conjunction of the axioms of real numbers 'reduce', then the knowledge in (i) arises from the knowledge that

(ii) $\Box(AX_{NBG} \supset R')$

together with the knowledge that

(iii) if $\Diamond R'$ then $\Diamond AX_{R}$;

(ii) is knowledge of necessity rather than of possibility,[14] and on the Fregean view, this is unproblematic; and (iii) is knowledge that results by the Conditional Possibility Principle just given.) But it is essential to this example that NBG be at least as rich as R. From the Fregean standpoint, any reason to believe that $\Diamond AX_{NBG}$ has to rest either on a reason to believe $\Diamond T$ for some *richer* theory T or else on a reason to believe AX_{NBG}. The deflationist cannot allow that there is any reason to believe either AX_{NBG} or any other mathematical theory. It is *compatible* with deflationism that there is an empirical theory T richer than NBG which can reasonably be believed; but (a) it is hard to believe that there is any plausible empirical theory in which NBG can be embedded, and (b) it is totally implausible that our reasons for believing that $\Diamond AX_{NBG}$ should rest entirely on reasons to believe any specific empirical theory. So from a Fregean standpoint, deflationism is simply not a viable position.

---

free variables introduced in the generalized substitution (which are unrestricted). It is easy to see that if a generalized substitution instance of B has a model, so does B itself, so the Conditional Possibility Principle is validated by the semantics of the appendix.

[13] This principal is valid as it stands on a free logic like that of Dana Scott's 'Existence and description in formal logic'. If one prefers a free logic like that of Tyler Burge's 'Truth and Singular terms', where each assertion of the form 'if $p(t_1, \ldots t_n)$ then $\exists x(x = t_1)$ & . . . & $\exists x(x = t_n)$' for atomic p is regarded as a truth of logic, then the principle must be modified slightly (say by redefining 'substitution instance' to mean a substitution instance in the normal sense conjoined with a clause of the form '$\exists x(x = t)$' for each term t in the sentence in which the substitutions are being made).

[14] Of course, knowledge of possibility is derivable from it: for example, that $\Diamond(AX_{NBG} \supset R')$, or that $\Diamond AX_{NBG} \supset \Diamond R'$; presumably, however, the Fregean view is that knowledge of possibility is never problematic if it is derivable from knowledge of necessity. (I count a knowledge claim as involving knowledge of possibility if its formulation contains at least one positive occurrence of '$\Diamond$' or at least one negative occurrence of '$\Box$'.)

I don't find this fact terribly upsetting, however, because I don't think that the Fregean viewpoint has a great deal of plausibility. In the first place, consider a point I made earlier, that part of our reason for believing that $\Diamond AX_{NBG}$ is the fact that we have been unable to derive any contradictions from AX_{NBG}. I argued then that this was a point that a deflationist could consistently recognize; I now want to observe that an advocate of *the Fregean position* could *not* recognize this point. For our inability to derive a contradiction from AX_{NBG} certainly doesn't give us reason to think that *actually* AX_{NBG}: if our reasons for believing that $\Diamond AX_{NBG}$ had to be based wholly on reasons for believing that actually AX_{NBG}, our inability to derive a contradiction from AX_{NBG} would be irrelevant to our knowledge that $\Diamond AX_{NBG}$.[15]

[15] It may seem that I am slurring over a complication. For it may seem that the fact that after persistent efforts we have not succeeded in deriving a contradiction from AX_{NBG} doesn't in itself provide evidence for $\Diamond AX_{NBG}$; rather, it provides evidence for a claim of *impossibility*, namely the impossibility of there being a derivation of a contradiction from AX_{NBG}. If this is right, then we need to establish a connection between this impossibility claim and the possibility claim $\Diamond AX_{NBG}$. Such a connection is established by the modal completeness theorem for first order logic, discussed in the next section.

How would this affect the argument in the text against the Fregean viewpoint? It might initially be thought to undercut the argument: for there is nothing in the Fregean viewpoint to rule out acceptance of the impossibility of deriving a contradiction from NBG on the basis of failures to find such derivations; and once that impossibility claim is accepted, it would appear that an application of the modal completeness theorem would yield knowledge that $\Diamond AX_{NBG}$. But of course the question is, how is the completeness theorem known? Knowledge of the modal completeness theorem is knowledge of possibility, and the proof of it sketched in section 4 assumes that $\Diamond AX_{NBG}$ (though doubtless we could make do with $\Diamond AX_M$ for a mathematical theory M somewhat weaker than NBG). Consequently, from a Fregean viewpoint it is hard to see how one could ever apply the theorem unless one already knew of an actual structure in which NBG (or the hypothetical weaker theory M) could be embedded. But that would mean that there would be no chance of adding to the credibility of the claim $\Diamond AX_{NBG}$ (or the claim $\Diamond AX_M$, if an appropriate weaker M is found) by persistently trying to derive a contradiction from NBG (or M) and consistently failing.

We see, then, that describing the epistemological situation as in the opening paragraph of this note would not help the Fregean. It would, though, create a problem for the non-Fregean (whether the non-Fregean be a platonist or a deflationist): for if our failures to derive a contradiction from AX_{NBG} only give reason to believe $\Diamond AX_{NBG}$ if we presuppose modal completeness, and that requires $\Diamond AX_{NBG}$, then such appeal to our failure to derive a contradiction is problematic from the non-Fregean position as well. Fortunately, however, the more complicated description of the epistemological situation seems wrong. It depends on thinking that the record of failures to find a contradiction could only enter into the epistemological picture as a premise to an enumerative induction. The right way to look at the matter, though, is as an inference to the best explanation: the assumption that $\Diamond AX_{NBG}$ is the most plausible explanation of the failure to find a contradiction in NBG. Modal completeness is irrelevant. (What is relevant is what I call in section 4 the 'weak modal soundness' of first order logic, but this is not something that one proves by $\Diamond AX_{NBG}$.)

In the second place, much of the motivation for the Fregean position is lost when we move from the crude formulation that Frege actually gives for his position (in the quotation above from §95 of *The Foundations of Arithmetic*) to the more defensible formulation that I have given. The motivation for the crude Fregean position is that it provides a simple solution to an epistemological problem, the problem of explaining the source of our knowledge of possibility. The crude Fregean position is that there really is no problem here: the source of our knowledge of possibility is just knowledge of actuality. The more defensible alteration of the Fregean position gives up this advantage: there is knowledge of possibility (not based solely on knowledge of necessity) that is not based on knowledge of actuality, but on 'reflection on relations of logical form'. The ability to 'reflect on relations of logical form' is supposed to allow us to know each instance of the schema '$\Diamond\psi \supset \Diamond\phi$' where ϕ is non-modal and ψ is a generalized substitution instance of it; but is it so clear that any motivated account of how we 'reflect on logical form' so as to come to this knowledge wouldn't also provide an account of how we know categorically some claims of form $\Diamond\phi$? After all, any claim of the same logical form as

(S) $\Diamond$(there are at least $10^{10^{10}}$ apples)

is also a logical truth. So why can't 'reflection on logical form' *show* that it is a logical truth? Why do we need to rest all our confidence in (S) on the claim that there actually are at least $10^{10^{10}}$ *somethings*? If we do need to rest knowledge of (S) on knowledge of actuality, that is rather surprising. What motivation is there for granting that we can have knowledge of possibility through 'reflection on logical form', but at the same time denying that knowledge of a simple possibility claim like (S) can be known by the same process?

Indeed, it seems to me that the Fregean position leads to quite counterintuitive consequences, for it seems clear to me that we have much more solid reason to believe (S) than to believe in the existence of $10^{10^{10}}$ entities of any kind. Certainly the claim that there are at least $10^{10^{10}}$ *physical* entities is not obvious. (I am inclined to believe it, for I am inclined to believe that regions of space are physical and that there are infinitely many of them; but I am much less confident of this than I am of (S).) And in my view, the claim that there are at least $10^{10^{10}}$ mathematical entities is far *less* obvious – in fact, I don't think there are *any* such entities – so I certainly wouldn't want to rest my belief in (S) on *that*. The idea that I should be as uncertain of (S) as I am of the claim that there are infinitely many physical or mathematical entities in the universe seems preposterous, and since the Fregean view has this consequence, I would need a much better motivation for that view before I could take it seriously.

The points I have made here for (S) arise for the claim '$\Diamond AX_{NBG}$' as well. That is, if this claim is true, so is every other claim of the same logical form; so why can't 'reflection on logical form' (whatever exactly that is) or whatever other process or combination of processes one needs to account for modal reasoning give us reason to believe it? Indeed, the example of '$\Diamond AX_{NBG}$' makes clear that the claim of even the crude Fregean view to be epistemologically pure was a hoax. The crude Fregean view was presented as having the epistemological advantage that it makes knowledge of possibility depend entirely on knowledge of actuality – but *mathematical* actuality is the only actuality that could work in the case of the claim '$\Diamond AX_{NBG}$', and once this is seen, it is hard to see how there is an epistemological advantage. The problem of how I know that it is *logically possible* that AX_{NBG} is 'solved' on the Fregean view by saying that I know that there *actually are* the entities that AX_{NBG} says there are and that they are interrelated as AX_{NBG} says. I find it hard to grasp how anyone could know (or have reason to believe) that such platonic entities actually exist (as opposed to being merely logically possible); or how anyone could know (or have reason to believe) that if such entities do actually exist, then they are related in one way rather than in some other. Consequently, the idea that I could explain my knowledge (or my reason to believe) that $\Diamond AX_{NBG}$ by reference to such knowledge of the platonic realm seems to me a total obfuscation of the real epistemological issues about knowledge of logical modalities.

3

Deflationism is, of course, a non-realist philosophy of mathematics: it holds that we cannot know (or have any reason to believe in) the existence of mathematical entities or the truth of ordinary mathematical claims; indeed, it would be natural to couple deflationism with the further claim that we have good reason to believe that there are no mathematical entities and hence that most ordinary mathematical claims are false.[16]

It has been widely held that the most serious difficulty facing any non-realist philosophy of mathematics is *the problem of application*: how can one account for the utility that reasoning about mathematical entities has for disciplines other than mathematics, if mathematics isn't construed in a realistic fashion?[17] Applied to deflationism, the problem

[16] Strictly speaking, the existentially quantified assertions will all be false and the universally quantified ones all vacuously true.

is: how can one explain the applicability of mathematics to disciplines other than mathematics, without assuming that ordinary mathematical claims (including those claims that assert the existence of mathematical entities) are true?

In a book I wrote several years ago I focused on one aspect of the problem of application: the problem of explaining the applicability of mathematics *to physical science* (and to everyday empirical reasoning), without assuming the truth of the mathematics that was being applied.[18] But there are other aspects to the problem of application, and a main task of the rest of this essay will be to say something about one of the more pressing aspects: the problem will be to explain the applicability of mathematics (in this case, proof theory and model theory) *to the study of logical reasoning*, without assuming the truth of the mathematics that is being applied.

Before turning to this main topic I want to say something about the applicability of mathematics to physical science and to ordinary empirical reasoning. Any account of the usefulness of a mathematical theory in dealing with the physical world will say that this usefulness depends on two things:

(a) the fact that the mathematical theory is 'mathematically good';
(b) the fact that the physical world is such as to make the mathematical theory particularly useful in describing it.

Different accounts of the usefulness of mathematics in application to the physical world will differ as to how (a) and (b) are to be elaborated.

A deflationist account of the application of mathematics must involve two claims. As regards (a), it must say that 'mathematical goodness' does not involve truth, but only something less demanding, such as consistency.[19] (This is strictly inaccurate, involving an inappropriate

[17] For instance, Frege says that 'it is applicability alone which elevates arithmetic from a game to the rank of a science' (Frege, *Grundgesetze der Arithmetik*, vol. II, section 92.) The point has been most thoroughly developed by Hilary Putnam in *The Philosophy of Logic.*

[18] *Science without Numbers.*

[19] Of course, a deflationist can and will recognize that not all consistent mathematical theories (and not all mathematical theories that are *strongly* consistent in the sense shortly to be defined) are of equal mathematical interest – just as the platonist can and will say that not all *true* mathematical theories are of equal mathematical interest. What makes a mathematical theory interesting is a complicated matter – richness in consequences is one factor, relevance to prior work in mathematics and in science is another, elegance is a third and doubtless there are further factors still. There is no need to discuss such factors here, for they are ones that the platonist and the deflationist can agree on. What are important in the present context are the features of mathematical goodness that go beyond

cont. over page

semantic ascent. The more accurate formulation is that the deflationist must claim that in explaining the application of a mathematical theory, we do not need to assume the conjunction of its axioms, since that conjunction isn't logically true; he or she must claim that instead we need to assume only something weaker which *is* logically true, such as the result of prefixing the conjunction of the mathematical axioms with the modal operator '$\Diamond$'. For the moment I will forgo such accuracy for the sake of naturalness.) As regards (b), the deflationist must be able to formulate the facts about the physical world that make the mathematical theory 'fit it', without assuming in the formulation any standard mathematics (either the mathematical theory whose usefulness is being explained or some other mathematical theory). For if we have to assume the truth of standard mathematics anywhere in our account – in (b) or in (a) – then a deflationist would have to hold that such an account was unknowable.

So a deflationist has two tasks corresponding to (a) and (b) above. The task corresponding to (b) is by far the more difficult, but since I have treated it at length in *Science without Numbers* and since it involves issues rather far removed from those of the rest of this essay, I will say no more about it here. The task corresponding to (a) is much easier. I have discussed it too in my book and in a more elementary paper;[20] but here I will need to summarize quickly what I said about it.

My conclusion was that a mathematical theory needn't be true to be good; and, indeed, if it were true, this wouldn't be enough for it to be good, for a good mathematical theory must have a property that might be called *strong consistency* or *conservativeness* and that doesn't follow from truth alone. To say that a mathematical theory M is strongly consistent is to say roughly that if you take any theory T that says nothing about mathematical entities, and add T to M, then if T is consistent, so is T + M.[21] Although strong consistency doesn't follow

---

interestingness. For the platonist, a mathematical theory should not only be interesting, it should be *true*; and my question is, what serves the role for the deflationist that truth serves for the platonist?

[20] Essay 2. For a discussion of some more technical aspects of the issues surrounding (a) (including replies to some technical objections that have been raised), see essay 4.

[21] More formally, for any theory T let T* be the result of restricting all variables of T by the condition 'is not mathematical'. Restrict attention to theories T that don't contain the predicate 'mathematical' and also don't contain any specifically mathematical vocabulary such as 'set'. Then a mathematical theory M is strongly consistent if for any such T, if T is consistent then so is T* + M. (The restriction on the vocabulary of T is needed to rule out T implying that there are mathematical objects, or that there are objects such as sets that M implies are mathematical.) In cases of interest, the intended ontology of T will include no mathematical objects. Still, if M is an impure mathematical

cont. over page

from truth, it does follow from *necessary* truth; I believe that the widespread view of mathematics as necessarily true shows an implicit recognition of the importance of strong consistency. Strong consistency, however, is weaker than necessary truth, for strongly consistent theories needn't be true at all. As for ordinary consistency, this is not in general a sufficient requirement on a mathematical theory. There is an important class of mathematical theories ('pure' mathematical theories – those dealing with mathematical entities alone) for which ordinary consistency entails strong consistency; so for such theories the only requirement is that they be consistent in the ordinary sense. But there is another important class of mathematical theories (e.g., certain versions of set theory with urelements which play an important role in the application of pure mathematical theories) for which ordinary consistency does not entail strong consistency; and for these theories strong consistency is the important notion. I have argued in the works mentioned that in explaining the application of mathematics to the physical world we never need assume that the mathematics is true, we need only assume that it is strongly consistent (i.e., conservative).

This sounds like an account of application that is congenial to the deflationist: the truth of standard mathematics needn't be assumed in the account, so there seems to be no problem in reconciling the application of mathematics with the deflationist's claim that the truth of standard mathematics can't be known.

---

theory, the domains of T* and M will overlap, and so will their non-logical vocabularies. For instance, if M is the most useful version of set theory with urelements, the domains of M and of T* will include all non-mathematical urelements; and all the vocabulary of T will appear in M too, in the comprehension schemata. (Similarly, if M is the most useful version of impure number theory, it will contain an operator 'the number of *x* such that F*x*', and through this the ontology and vocabulary of T will appear in the mathematics. For instance, one of the instances of the induction schema will be that if there is a planet with 0 moons, and if for each n such that there is a planet with n moons there is also a planet with n + 1 moons, then for all n there is a planet with n moons.) The fact that the ontology and vocabulary of T and M overlap explains why strong consistency doesn't reduce to ordinary consistency for typical impure mathematical theories.

The above definition of strong consistency is essentially the one given in *Science without Numbers*. (In the book I followed the standard artifice of regarding logic as ruling out the empty domain; because of this, I added some extra complexity in the definition of strong consistency, but there is no point in adding it here since the artifice has been dropped. I also misformulated the restriction on the vocabulary of T in fn. 8, though it should have been clear from the context that this was what was meant.) In the book I did not contemplate adding mathematics to theories that contained modal operators, for I was concerned only with adding them to physical theories and I did not (and do not) want to allow modal operators into nominalized physics. If we do consider the addition of mathematics to a theory T with modal operators, we might want to complicate the definition of T*; but there is no need to go into that here.

I think that this point is correct in spirit, but there is a difficulty that must be faced. For I have defined strong consistency in terms of ordinary consistency, and ordinary consistency is usually defined in terms of the existence of models. So won't the assertion that a theory is strongly consistent be an assertion about the existence of models? If so, then even though strong consistency doesn't entail truth, it is still hard to see how a deflationist could ever claim to know any theory to be strongly consistent. And if the deflationist can't claim to know that, it is certainly awkward for him or her to maintain that the strong consistency of a theory is essential to its application.

Many people have objected along these lines to the account of mathematics put forward in my book, and with considerable justification.[22] But from what I have said so far in this paper, it should be clear in outline how I now want to handle the objection. I want to say that in explaining the application of a mathematical theory M to the physical world, it is not strictly accurate to say that we need to assume the strong consistency of the mathematical theory. Rather, what we must assume is a certain modal claim, one which bears the same relation to the claim that M is strongly consistent that $\Diamond AX_M$ bears to the claim that M is consistent in the ordinary sense. The points I made several paragraphs back about the relation between strong consistency, ordinary consistency, truth and necessary truth should really have been made at the object level: instead of saying that strong consistency is entailed by necessary truth and entails consistency, but neither entails nor is entailed by truth, I should have said that the claim Q that is the modal analogue of the strong consistency of M is entailed by $\Box AX_M$, entails $\Diamond AX_M$ and neither entails nor is entailed by AX_M. Similarly, for the point about explaining the utility of M: instead of saying that we don't need to assume the truth of M, but only its strong consistency, I should have said that we don't need to assume AX_M but only Q.

That, it should be clear, is how I want to handle the objection. But *can* I handle it in this way? There is a technical difficulty to doing so, for there is a technical difficulty in figuring out exactly how Q (the modal analogue of strong consistency) is to be formulated.

Before turning to this technical difficulty, I want to discuss an earlier technical difficulty that I mentioned in section 1: the difficulty about

[22] For example, Michael Resnik and David Malament in their reviews of my book; Charles Chihara in 'A simple type theory without platonic domains' and Michael Detlefsen in *Hilbert's Program: an essay on mathematical instrumentalism*. (For some reason Malament confines his objection to the case where a non-axiomatizable logic is at issue, but in fact it applies equally well to axiomatizable logics, since formal derivations are just as suspect from a nominalist viewpoint as models are.)

theories that are not finitely axiomatized. In section 1, I suggested that 'knowledge of the consistency of the theory of linear order' is really just modal knowledge: it is knowledge of the form ◇B, where B abbreviates the conjunction of the axioms of the theory of linear order. But suppose we consider a theory that is not finitely axiomatizable. Can't we know that such a theory is consistent? And how do we represent such knowledge as modal knowledge given our inability to conjoin all of the infinitely many axioms?

There is more than one way to respond to this objection; my current inclination is to respond by introducing a further logical device that will allow us to finitely axiomatize theories which, without the device, can't be so axiomatized. The device I have in mind is what is called a 'substitutional quantifier'. The name seems to me misleading: it is not a quantifier at all, as quantifiers are normally understood; rather, it is simply a device for representing sufficiently regular infinite conjunctions in a finite notation.[23]

The non-finitely axiomatized theories we ordinarily use are theories with quite regular infinite collections of axioms. For the theories consist of finitely many axioms plus finitely many axiom *schemata*: schemata like 'For every formula F, '$\exists z \forall y\, (y \in z \equiv F)$' is to be an axiom.' To represent this finitely, we merely need to conjoin all the infinitely many sentences in the language of the form '$\exists z \forall y\, (y \in z \equiv F)$'; and we can do that if we have a substitutional quantifier 'Π' with formulas as the substitution class, for we simply say '$\Pi F \exists z \forall y\, (y \in z \equiv F)$.' (This is not a metalinguistic assertion but an infinitary conjunction of non-metalinguistic claims: it is no more metalinguistic than is a finite conjunction like '$\exists z \forall y$ ($y \in z \equiv y$ is a cat) & $\exists z \forall y$ ($y \in z \equiv y$ is a dog).') Anything that we normally regard as a mathematical theory can be finitely axiomatized using such a device, so any knowledge that we possess of the consistency of mathematical theories can be represented in the form ◇B if we're allowed to use a substitutional quantifier in formulating B.

One advantage of this way of solving the technical difficulty about non-finitely axiomatized theories is that it solves the technical difficulty about the modal analogue of strong consistency as well. Strong consistency is defined in terms of ordinary consistency as follows:

(iii) for any theory T, if T is consistent, then so is $T^* + M$;

here T^* is the result of restricting all the quantifiers of T to non-mathematical entities. What is the modal analogue of this? It will differ

[23] The view of substitutional quantification implicit in these remarks is set out in more detail in my review of Dale Gottlieb's book *Ontological Economy: substitutional quantification and mathematics*.

from (iii) in not being metalinguistic: instead of saying that if T is consistent, then so is $T^* + M$, we will say that if $\Diamond AX_T$, then $\Diamond((AX_T)^*$ & $AX_M)$. But then, how do we *generalize* such an object-level statement to all theories T? Again we must invoke a substitutional quantifier (or some other form of infinite conjunction); we must say

(iii*) ΠB (if $\Diamond B$, then $\Diamond(B^*$ & $AX_M))$

My solution to the technical difficulty raised several pages back, then, is that 'knowledge of conservativeness' is really just modal knowledge of the form (iii*). Such modal knowledge does not require knowledge that AX_M (much less that $\Box AX_M$); and this is all to the good, since AX_M (and $\Box AX_M$) entails the existence of mathematical entities and consequently is not logically true. For certain mathematical theories, (iii*) will be stronger than the claim $\Diamond AX_M$, but even so, it is a logical truth. And the application of the theory M to the physical world requires only the logical truth (iii*), it does not require a claim like AX_M which asserts the existence of mathematical entities and hence is not logically true.

4

So far I've argued that for lots of purposes where we might seem to need notions like consequence and consistency and strong consistency or conservativeness, we really need only modal analogues of these notions, and that this is good because you can explain how facts involving the modal analogues are known without postulating knowledge of mathematical entities. In other words, in many contexts metalogical notions (at least notions of *semantic* metalogic) are dispensable in favour of corresponding object-level notions.

But an important problem remains for a deflationist: the problem of accounting for the utility of reasoning at the metalogical level rather than at the object level. The problem is especially striking for proof-theoretic reasoning, since here there seems to be no object-level analogue. An object-level assertion is one that makes no reference to sentences or formulas (or abstract analogues of them such as propositions); consequently, it can make no reference to axioms or rules of inference or formal derivations. It is hard to see how any such assertion could in any interesting sense be an analogue of an assertion that one sentence is (or is not) formally derivable from another, using a given system of logical axioms and logical inference rules.

How then are we to account for the utility of proof-theoretic reasoning? Traditionally, proof-theoretic concepts are defined in terms of mathematical entities, with the result that proof-theoretic reasoning becomes reasoning about mathematical entities. If we accept the usual definitions of proof-theoretic concepts, then a deflationist cannot regard proof theory as a subject of which we can have any knowledge. So how can a deflationist account for its utility?

There seem to be two ways for the deflationist to try to solve this problem. The first is to reject the usual definition of proof-theoretic concepts, and provide alternative definitions which make no reference to mathematical entities. The idea would be to show that if proof-theoretic notions such as formal derivability are understood in terms of these alternative definitions, then claims about formal derivability (etc.) can be known consistently with deflationism, that is, consistently with there being no knowledge of mathematical entities.

One way to try to work out this first approach would be to take derivability to be some sort of modal notion. First we could try to get a sufficiently powerful theory of actual inscriptions, without introducing modality: in terms of such a theory, we could explain notions like 'e is a well-formed inscription', 'e and f are type-identical inscriptions', 'd is (an inscription that constitutes) a derivation (according to system F)', and various predicates of inscriptions that describe them structurally ('being an A-inscription', where A is an expression type). I believe that this part of the project could be worked out in first order logic (though some care is needed because there are no means here to make recursive definitions explicit). The second part of the project would be to make some modal extension: in it, we might hope, we could understand 'A is derivable' to mean 'it is possible that there is a derivation whose last line is an A-inscription', instead of (as the platonist would have it) as meaning that there actually exists a certain type of abstract sequence of abstract analogues of the symbols. This has considerable initial attraction as an account of the ordinary meaning of 'derivable'. I am, though, reluctant to introduce a new type of possibility beyond strictly logical possibility here, unless we can define it from strictly logical possibility plus other acceptable notions; and there are substantial difficulties that must be overcome for the project of doing this.[24] We could avoid these

[24] For instance, there are at least two obstacles to taking the relevant sort of possibility to be consistency with the first order theory that one obtained in the first part of the project. The first obstacle arises from the fact that logical possibility is thoroughly anti-essentialist: this poses a problem for translating sentences in which 'derivable' occurs in the scope of a quantifier. ('He uttered an underivable inscription' would always be false on the naive translation.) The natural way to solve this problem would be to take all

cont. over page

difficulties with additional logical devices like a substitutional quantifier, but these might make the appeal to modality unnecessary anyway. I will not pursue these matters here.

In any case, carrying out this programme wouldn't really solve the more general problem raised two paragraphs back. The problem was for the deflationist to account for the utility of proof theory without assuming the truth of mathematics. And in presenting the problem it was assumed that proof theory meant *platonistic* proof theory. The first deflationist approach to this problem says that there is a nominalistic proof theory that is just as good as platonistic proof theory, and that the nominalist has no difficulty in accounting for the utility of *that*. But unless more is said, it looks as if this is merely changing the subject from the original question, which was how the utility of the platonistic theory is to be explained.

My primary interest then will be with the second deflationist approach to the problem of explaining the utility of proof theory: the approach of trying to explain the utility of platonistic proof theory without assuming it true. I will explain shortly how this second approach can be carried out.

But first I want to shift attention from proof theory back to semantics. The problem with which I began this section – the problem of accounting for the utility of mathematics in metalogical reasoning – is a problem that arises for semantics as well as for proof theory (though it may initially seem less striking a problem in the case of semantics, since there we have object-level analogues of our metalogical notions). And again, there are two different approaches that one might be inclined to take.

The first approach would be to reject the usual model-theoretic definitions of semantic consequence and similar notions, and propose alternative definitions instead. Can this be done? In a sense it can. We can say

> (5) B is a semantic consequence of Γ (where Γ is a finite list of sentences together with a finite list of schemata) if and only if □ (if all members of Γ are true, then B is true).

'quantification in' to be substitutional. The second obstacle is that the consistency with axiomatic proof theory (even axiomatic platonistic proof theory) of the existence of a proof is not sufficient for provability in the normal sense: incompleteness theorems give cases of unprovable formulas where the assertion that there is a proof is consistent with proof theory. The natural way to solve this would be to say that for A to be provable, the existence of a proof of A must be compatible with a (nominalistic or platonistic) proof theory stated in a powerful logic that can rule out derivations that are not genuinely finite: say a logic with the quantifier 'there are only finitely many', or a logic with a substitutional quantifier or other device of infinite conjunction.

This definition would be objectionable if the word 'true' here were used in a 'transcendent' sense, that is, in a sense in which 'Snow is white' wouldn't have been true if 'white' had meant 'green'; for in that transcendent sense the modal claim on the right of (5) is going to be false even if B is a consequence of Γ in the usual sense. But let us understand 'true' instead in the 'immanent' sense,[25] that is, as applicable most directly to one's own language and as obeying there the principle that □('Snow is white' is true if and only if snow is white). We can define such an immanent sense of 'true' using substitutional quantifiers: S is true if and only if $\Pi p(S = \text{'}p\text{'} \supset p)$. Or alternatively, one can follow Grover, Camp and Belnap and regard 'true' (or at least 'true' in the immanent sense) not as a predicate at all but (roughly speaking) as simply the means by which substitutional quantifiers with sentences as substituends are represented in English.[26] In either case, it turns out that (5) is equivalent within standard mathematics to □(if AX_{Γ} then B), where AX_{Γ} is a conjunction of all the axioms in Γ (using substitutional quantifiers to conjoin the instances of the schemata). On this alternative to the usual way of defining consequence, what I earlier called the modal analogue of a claim that one thing was a consequence of another wouldn't really be an *analogue* at all, it would simply be what the consequence claim *means*.

I don't attach a great deal of philosophical significance to the possibility of defining semantic consequence in this nonstandard way. My reason for mentioning it is only to point out that even if it is adopted, it leaves an important problem unsolved: namely, how is a deflationist going to explain the utility of *model-theoretic* definitions of semantic consequence? Even assuming that it is somehow better to define consequence modally in the way just indicated (a claim on which I take no stand), still model-theoretic semantics has proved enormously useful; and it is not immediately evident *why* it should be useful if consequence is really to be defined modally via an immanent notion of truth, or if consequence claims are to be rejected as strictly speaking unknowable and only modal analogues of them are to be claimed as knowable. So a deflationist must give an account of why standard uses of model theory are legitimate even if model theory isn't true, just as he or she must give an account of why standard uses of platonistic proof theory are legitimate even if platonistic proof theory isn't true.

In order to provide such accounts, we must first ask to what uses model-theoretic semantics and proof theory are standardly put; only

25 This use of the terms 'transcendent' and 'immanent' was suggested by (but isn't quite the same as) Quine's use of these terms in *Philosophy of Logic*.

26 'A prosentential theory of truth'.

then can we ask whether these uses are explainable from a deflationist viewpoint. I will not attempt an exhaustive account of the uses to which model-theoretic semantics and proof theory are put, but it seems to me that the central uses are as devices for finding out about logical possibility. Model theory is used for this purpose via the instances of the following two schemata: the *model-theoretic possibility schema*

(MTP) If there is a model for 'A' then $\Diamond$A;

and the *model existence schema*

(ME) If there is no model for 'A' then -$\Diamond$A.[27]

And proof theory is used for this purpose via the *modal soundness schema*

(MS) If there is a proof of '-A' in F then -$\Diamond$A,

which holds for any reasonable formal system F for any fragment of logic; and via the *modal completeness schema*

(MC) If there is no proof of '-A' in F then $\Diamond$A,

which holds for certain areas of logic (i.e., certain types of sentence A) and certain sufficiently strong formal systems for those areas of logic.

From the platonist standpoint, all of the instances of these four schemata are true (for the appropriate formal systems F, in the case of the last two schemata).[28]

What *justification* might a platonist offer for these schemata? I'll focus on MTP and MS, since these seem the evidentially primary ones. (The

[27] If A contains free variables, interpret 'model' in (MTP) and (ME) to mean a model together with an assignment function for the free variables.

[28] Actually there is some question as to whether we should expect (ME) to hold for arbitrary logics. Even for first order logic, it seems somewhat surprising that (ME) holds (in the same way that the 'Skolem paradox' seems somewhat surprising): just as the universe of classes is too big to form a countable model, it is too big to form a class and hence too big to form any model at all; so just as it seems somewhat surprising that the sentence AX_{NBG} that formulates the theory of this universe of classes should have a countable model, it seems a bit surprising that it should have a model at all. But the classical completeness theorem shows that it does have a model if it is not formally inconsistent; this makes (ME) a set-theoretic consequence of (MS) in the case of first order logic, which surely makes it platonistically acceptable in the case of first order logic. Still, it may not be acceptable for arbitrary logics. It seems clear that if it fails for some logic – for example, if there is a sentence A formulatable in that logic which expresses enough about the 'intended model' of NBG to preclude there being any class that can be a model of A – then the usual model-theoretic definition of consistency should be regarded as extensionally incorrect for that logic: the imagined sentence A is intuitively consistent, even if it has no model.

other two follow from these, of course, by the classical completeness theorem.) In the case of MTP, each first order instance (instance where the instantiating formula is non-modal) is almost immediate from an instance of the Conditional Possibility Principle of section 2 (together with NBG and an instance of '$A \supset \Diamond A$').[29] (I'll discuss the case where the instantiating formula is modal later.) In the case of MS, though, the situation is more delicate. To simplify the discussion, we'll pick an F in which in any application of a rule of inference, the formula being inferred is a logical consequence of the formula from which it is inferred.[30] A platonist might be tempted to argue for the validity of MS (as applied to F) by induction. More accurately, we could argue informally that each instance of

(MS′) If there is a proof of 'B' in F then $\Box B$

is true, by induction on the length of the proof: certainly '$\Box B$' is true when B is a logical axiom of any reasonable F, and if B is directly inferable from $B_1, \ldots, B_n$ then '$\Box(B_1 \,\&\, \ldots \,\&\, B_n \supset B)$' is true, and so using the induction hypothesis '$\Box B$' is true. And of course each instance of MS follows from an instance of MS′.

But this informal induction goes beyond the modal consequences of standard mathematics, if 'standard mathematics' is taken in the normal way: for in the induction I have utilized a notion of truth (for sentences in the modal language) that has not been defined. Indeed, besides the usual problem with defining truth, and hence carrying out the induction, that the Tarski indefinability theorem poses, there is a further problem: what would the recursion clause for formulas that begin with a logical possibility operator be? In the appendix, in defining truth *in a model*, I use the clause

'$\Diamond A$' is true *in model M* $\equiv$ there is a model in which A is true.

[29] For any formula B containing predicates $p_1, \ldots, p_k$, let $M, E_1, \ldots, E_k$ be variables not in B, and let $\hat{B}$ be the generalized substitution instance that results from B by restricting quantifiers and free variables by the predicate '*x* is in M' and by replacing (for each i and each $v_1, \ldots, v_{n_i}$) '$p_i(v_1, \ldots, v_{n_i})$' by '$<v_1, \ldots, v_{n_i}>$ is in E_i'. Also, let B^+ be the existential closure of $\hat{B}$. (Example: if B is '$\forall y\,(p_1(y,z))$', $\hat{B}$ is '$z \in M \,\&\, (\forall y \in M)(<y,z> \in E_1)$', and B^+ is '$\exists M \exists E_1 \exists z[z \in M \,\&\, (\forall y \in M)(<y,z> \in E_1)]$'.) $\Diamond\hat{B}$ is equivalent to $\Diamond B^+$; this plus the Conditional Possibility Principle yields $\Diamond B^+ \supset \Diamond B$. So to get the instance of MTP we need only argue in NBG for 'If there is a model M and an assignment function s in which B comes out true, then $\Diamond B^+$.' But in fact we can inductively argue in NBG that if there is such an M and s then *actually* B^+; so since possibility follows from actuality, we have the instance of MTP.

[30] In the case of non-modal quantification theory, the system of Hunter's *Metalogic* has this property: instead of containing a rule of universal generalization it contains the principle that the universal generalizations of quantificational axioms are quantificational axioms. The analogous strategy can be used in modal systems: avoid a necessitation rule by taking necessitations of all axioms to be axioms.

(For perspicuity I consider only the case where A is a sentence.) It might be natural to say by analogy

(*) '◇A' is true ≡ there is a model in which A is true.

But if 'standard mathematics' is taken as a first order axiomatic theory, this won't do very well as part of a recursive definition of truth, for then '◇A' together with standard mathematics won't in general entail '"◇A" is true.'[31] Reason: by Gödel's second incompleteness theorem, there are models of axiomatic set theory NBG in which the sentence 'NBG has a model' will be false. In such a model, '◇NBG' will be true (it's true in all models), but on the proposed definition of truth '"◇NBG" is true' will be false in the model.

I am not denying, of course, that a platonist should accept the material equivalence (*) (for sentences A not containing terms like 'true' – see footnote 31). Indeed, assuming that '"A" is true' is equivalent to 'A' (for sentences not containing 'true'), (*) is simply the conjunction of MTP and ME, which I have said a platonist ought to accept. The argument is, though, that there are models of set theory in which the left to right direction of (*) fails, so that *barring independent support of MS and hence of ME*, (*) does not have the kind of necessity we would like in a recursion clause of a definition of truth.

We might of course try to avoid this difficulty by first arguing for MS for non-modal formulas A, where a notion of the truth of modal formulas is not required; then using the result there (together with MTP there) to extend MS to modal formulas without iterated modalities, and so on. But there is still the difficulty that we can't provide the induction even for the case of non-modal formulas, by the Tarski indefinability theorem.

Indeed, the Gödelian example shows not only that there are problems in formalizing the intuitive proof of MS modally within standard mathematics: it shows that MS is not a consequence of standard mathematics alone, even for non-modal A and even in the strong modal logic given in the appendix. Let CON_{NBG} be the statement that there is no F-proof of not-NBG (where F is a proof procedure for first order logic). Then an instance of MS – indeed, an instance with a purely first order sentence as the substituend for the schematic letter 'A' – is

If -CON_{NBG} then -◇NBG;

equivalently,

(**) If ◇NBG then CON_{NBG}.

[31] The sentences A involved here don't contain 'true' or related terms, so the lessons of the semantic paradoxes do not do anything to make this conclusion palatable.

By Gödel's second incompleteness theorem, CON_{NBG} is not derivable from NBG. So there is a first order model in which NBG holds and CON_{NBG} doesn't. But the model theory in the appendix takes first order models to be modal models as well, so NBG doesn't imply CON_{NBG} modally either. But NBG does imply ◇NBG modally, so (**) cannot be a modal consequence of NBG.[32]

At this point in the original version of this essay I introduced a nonstandard sense of 'standard mathematics', employing substitutional quantifiers and an ω-rule governing them, from which MS does follow modally by an inductive proof like the one recently sketched (and in which (*) becomes a reasonable recursion clause for truth). It now seems to me, though, that this made the presentation of several points confusing, and the weight put on substitutional quantification may seem suspicious. It seems simpler just to say that a platonist just accepts MS, even though it is not provable in standard mathematics from more elementary modal principles. It is worth remarking (though how important this is I'm not sure) that in the modal logic of the appendix, the non-modal instances of MS are enough to generate (in standard mathematics in the usual sense) all instances of MTP and even of ME, including those where modal sentences are the substituends.[33]

I have been discussing the epistemological status, from a platonist perspective, of MTP and MS and the other lettered schemata. It is clear that once these schemata are available, the platonist can use ordinary

[32] In fact, of course, even the weakening of MS that drops the possibility operator is not a consequence of standard mathematics.

[33] For both MTP and ME, one uses an induction on the modal degree of the formula A that is the substituend (generalizing MTP and ME to formulas as in fn. 27). If A is of modal degree n + 1, let A* be the result of taking each sub-formula of form ◇ B for which B is degree 0 (i.e., non-modal) and replacing it by 'B v -B' if B has a model and by 'B & -B' otherwise. Now we've proved the degree 0 instances of MTP, and the degree 0 instances of ME follow by completeness from the degree 0 instances of MS, which we're assuming. Using these, we get (in S5) that if B non-modal then □{(◇B) ≡ (◇B)*}, and by a subinduction on the depth of the embedding of ◇B in A, □{A ≡ A*}. Consequently, (1) ◇A ≡ ◇(A*). Also, another trivial subinduction shows that any model is a model of A if and only if it is a model of A*, and consequently (2) A has a model if and only if A* has a model. But (1) and (2) and MTP for A* (which is of modal degree n) yield MTP for A; and analogously for ME.

Incidentally, the reliance on the degree 0 instances of MS is essential, even for MTP: for instance, if ◇NBG but NBG has no model, then -◇NBG has a model (indeed, it is true in *any* model), but -◇-◇NBG by the S5 axiom, so we have a degree 1 violation of MTP. Still, a given instance of MTP, with A the substituend formula, can be proved from NBG plus n degree 0 instances of MS, where n is the number of occurrences of logical operators in A. It is easy to calculate what the required instances of MS are (inspection of the proof above shows how to do it), and in a typical case most of them will be provable in NBG anyway so they won't really need to be added.

platonist model theory and proof theory for finding out about possibility and impossibility. But how is this of any help to a deflationist, who denies that the existence of mathematical entities and the truth of mathematical theories can be known? Assume that we construe 'proof' and 'model' in the usual platonist way. Then if there are no mathematical entities, ME and MC are simply invalid: their antecedents are true irrespective of whether ◇A or -◇A. And if there are no mathematical entities, MTP and MS are only *vacuously* true: they are useless as an aid to finding out about possibility and impossibility, because their antecedents can never be fulfilled. Admittedly, one could do a bit better by considering *concrete* proofs (made by actual physical inscriptions) and *concrete* models. By construing 'model' concretely, we'd get a 'weak MTP' that is non-vacuous (it would be little more than a restatement of the Conditional Possibility Principle, of course); but it would generate the logical possibility of rich structures only from controversial empirical premises. The 'weak MS' obtained by construing 'proof' concretely would be even more severely limited, since rigorous proofs in formal systems are rarely given for anything that is in the least complicated. Apparently, then, the deflationist has a problem.

In fact, though, the problem is easily solved. Instead of MTP and MS (or perhaps, in addition to their weak versions), the deflationist employs the following modal surrogates:

(MTP#) If □(NBG ⊃ there is a model for 'A') then ◇A

and

(MS#) If □(NBG ⊃ there is a proof of '-A' in F) then -◇A.

From these and the classical completeness theorem, one can derive

(ME#) If □(NBG ⊃ there is no model for 'A') then -◇A

and

(MC#) If □ (NBG ⊃ there is no proof of '-A' in F) then ◇A,

in the case where first order sentences are the only substituends; indeed, ME# can be argued to hold also where modal sentences are substituends.[34] The deflationist can use the hatched schemata in pretty much the same way the platonist used the unhatched ones: to find out that A is, or is not, logically consistent, it suffices to derive a model-theoretic or proof-theoretic statement from standard mathematics.

Could it be claimed that the deflationist has less reason to believe MTP# and MS# than the platonist has to believe MTP and MS? I do

[34] This will be clear from the remarks on MTP# below, in conjunction with fn. 33.

not see how this could be made plausible. First let's consider the instances of MTP# in which the instantiating formula is non-modal. Earlier (footnote 29) I presented a platonist argument for the corresponding instances of MTP, within standard mathematics; in effect, then, I showed that for non-modal A

$$\Box(\text{If NBG \& there is a model for `A', then } \Diamond\text{A.})$$

From this it follows (in S4) that

$$(\text{MTP}^*)\ \text{If } \Diamond(\text{NBG \& there is a model for `A'}), \text{ then } \Diamond\text{A}.$$

From this and the assumption ◇NBG, one gets the instance of MTP#. So the platonist's argument from NBG to MTP yields an argument from ◇NBG to MTP#. The deflationist does of course need to assume ◇NBG in order to accept MTP#, but if my earlier arguments in section 2 are correct, such consistency claims are ones that a deflationist can have perfectly good reason to believe. In any case, the deflationist's epistemological burden is strictly weaker than that of the platonist, who must believe not only that ◇NBG but that *actually* NBG.

The above reasoning can be extended to instances of MTP# where the instantiating formula is modal. In footnote 33 I presented a platonist argument for an arbitrary instance of MTP: the argument was from NBG plus a few instances of MS (which ones being easily calculable from the substituend sentence in MTP). Letting NBG_A be NBG with these added instances, the reasoning of the previous paragraph shows that the platonist's argument from NBG_A to the instance of MTP yields a deflationist argument from $\Diamond NBG_A$ to MTP#. (Actually to a slight strengthening of MTP#, one obtained by replacing 'NBG' in it by 'NBG_A'.) Again, the claim that the deflationist needs, $\Diamond NBG_A$, is a logical consequence of the one that the platonist needs, and I don't see how it can be plausibly argued that the deflationist is in worse epistemological shape than the platonist here.

What about the epistemological status of MS#? In the case of MS, the reader will recall, the platonist had no hope of rigorously deriving it from NBG (even in the strong modal logic of the appendix); nonetheless, MS is a claim that a platonist ought to accept as a primitive modal assumption. The deflationist, similarly, can accept MS# as a primitive modal assumption. Alternatively, the deflationist can derive each instance of MS# from *the possibility of* what the platonist assumes: i.e., from ◇(NBG + the corresponding instance of MS). (This is a trivial derivation in S5.) Either way, the deflationist seems to be in pretty good epistemological shape.

A platonist might respond to this by saying that in the case of MS we have an informal (and unformalizable) inductive argument in its favour; but that the deflationist has no such inductive argument in favour of sentences of form $\Diamond$(NBG + B) where B is an instance of MS, and so must ride piggyback on a platonist argument that he or she does not accept. I think, though, that this is wrong. To say that one accepts an informal inductive argument that cannot be formalized in one's theory is to say in effect that one accepts a stronger theory. In the case of the argument for MS, it might plausibly be argued that we are implicitly employing ordinary set theory to which a truth predicate has been added. (It is a truth predicate for sentences of set theory that don't contain it; and it is allowed to occur in the separation and replacement schemata.) In such a more powerful set theory, of course, MS is actually derivable, provided 'proof' in it is understood as 'proof not employing the new truth predicate'. But now, if the platonist can appeal to such a powerful theory S, the deflationist can appeal to $\Diamond AX_S$.[35] And since the instances of MS follow from S, the instances of $MS^{\#}$ follow from $\Diamond AX_S$, by the same argument as for $MTP^{\#}$.

Another alternative, for the platonist who takes the intuitive inductive argument for MS seriously, is to employ some sort of ω-rule. That is, the intuitive inductive argument shows that in ordinary set theory, one gets as theorems each instance of

(MS^k) If there is a proof of '-A' of length k, then -$\Diamond$A

(where k is any numeral and A any formula). The trouble is that ordinary set theory is ω-incomplete: one cannot get from these to 'For all n, if there is a proof of "-A" of length n then -$\Diamond$A.' The intuitive induction works by assuming an ω-rule that licenses this generalization from the various MS^k, or from the argument for them. Now, if this is the logic that the platonist employs, my reply of course is that the deflationist is allowed to employ the same logic. But if so, then the same sort of argument as above gets us from the platonist's 'derivation' of MS in the expanded logic to a deflationists 'derivation' of $MS^{\#}$ in the expanded logic. However the platonist twists and turns in an effort to avoid taking MS as simply a primitive assumption, the deflationist can twist and turn too.

I conclude that the deflationist has no more difficulty in using platonistic model theory and proof theory in finding out about possibility and impossibility than does the platonist.

In section 3, I distinguished two 'problems of application': the problem of application of mathematics to the physical world, and the

[35] S won't be finitely axiomatized; but we can avoid the use of substitutional quantifiers here, by using the theory consisting of the possibilization of each finite conjunction, as discussed in the postscript.

problem of application of mathematics to the study of logical reasoning. I have just outlined a solution to the latter problem; but readers may be puzzled by a disanalogy between this solution and the solution offered in *Science without Numbers* to the problem of application to the physical world. In solving both problems of application, I tried to legitimize a certain kind of instrumentalism about mathematics: I tried to argue that platonistic physics and platonistic metalogic were usable even if not true. But in *explaining* why the usability of these theories didn't depend on their truth, I had to do more work in the case of platonistic physics than in the case of platonistic metalogic. That is, the explanation given in my book for the legitimate usability of platonistic physics turns on the existence of a nominalistic physics. But the explanation I have just given for the legitimate usability of platonistic proof theory doesn't require the existence of a nominalistic proof theory. (I have stated that such a nominalistic proof theory may be possible, but my explanation of the usability of platonistic proof theory in finding out about possibility and impossibility did not turn on this.) What accounts for the difference?

The answer is that physics has an explanatory function: you need physical theories to explain physical phenomena. According to the form of nominalism I accept, one should not junk a platonistic explanation of a phenomenon unless there is a satisfactory nominalistic explanation to take its place. It is because of this principle that a satisfactory nominalistic formulation of physical theories is required. Now, the main role of platonistic proof theory is not explanatory. If I use proof theory as an aid to discovering whether B follows from A, it is not because the proof-theoretic principles are in any way explanatory of the fact that $\Box(A \supset B)$ or $-\Box(A \supset B)$; the proof theory is solely an instrument of discovery, and needn't be replaced by some other theory about which we must take a non-instrumentalist attitude.

I have not said that proof theory has *no* explanatory function, but only that its *central* function is not explanatory. There is a sense in which proof theory can be used to explain. Suppose I want to explain the historical fact that no one has ever produced a physical inscription which constitutes a formal derivation, in Kleene's system of logic, of an explicit contradiction. Intuitively, the reason that no one has ever produced such an inscription is that it is impossible that there be one; and here I think 'impossible' can be taken to mean 'logically inconsistent with certain assumptions we make about physical inscriptions'. Any codification of those assumptions is a theory of physical inscriptions, and from it we can explain the historical fact in question. One way to codify those assumptions about physical inscriptions is to formulate platonist proof theory and then add a principle saying that there is a certain kind of homomorphism mapping physical inscriptions into expressions in the platonist sense. This codification produces a platonist explanation of the historical fact in question. I also think that assumptions about physical inscriptions which are adequate to the purposes at hand can be stated nominalistically (without using devices going beyond first order logic, in this case); if so, then we will have a nominalistic explanation of the historical fact. I take it, though, that the use of proof theory to explain such historical facts is of less importance than its use in finding out about possibility and impossibility; and in the latter use, the platonist proof theory does not serve an explanatory function, and so no nominalistic proof theory is required.

I have been discussing a disanalogy between my treatments of physics and of proof theory; how does model theory enter into the picture? That is, in explaining the legitimate usability of platonistic model theory, did we (as with platonistic physics) need to develop a nominalistic analogue of the platonistic theory? Or (as with the central applications of platonistic proof theory) did we not? This question is largely verbal: it depends on whether we regard modal logic itself as an analogue of platonistic model theory. If we do so regard it, then model theory is like physics, and the earlier sections of this paper were devoted largely to developing the nominalistic analogue as a necessary prelude to explaining the legitimate usability of platonistic model theory.[36] If we do not so regard it, then model theory is like proof theory in its most central applications: we did not need a nominalistic analogue of model theory because model theory doesn't serve to explain anything, but simply serves as a tool for enabling us to find out more easily about possibility and impossibility. The difference between these two viewpoints is merely a difference between ways of looking at what has been done: whatever the viewpoint, the argument earlier in this section shows how one can explain the legitimate usability of platonistic model theory without assuming its truth.

5

In this essay I have been advocating the view that all mathematical knowledge that isn't straightforwardly empirical is knowledge of a purely logical sort. By 'mathematical knowledge' here I do not mean knowledge of the claims of mathematics. According to the view I have advocated (deflationism) there *is* no mathematical knowledge in that sense. Rather, by 'mathematical knowledge' I mean the sort of knowledge that those

[36] The prelude is necessary, for as I argued in essay 2, there is no way to explain the legitimate usability of metalogic by conservativeness alone if the underlying logic is taken to be non-modal.

who know a lot of mathematics have a lot of and those who know little mathematics have little of. If deflationism is false, this will include knowledge of mathematical claims; but whether deflationism is true or false, it will include a lot of knowledge that isn't knowledge of mathematical claims.

As hinted several times, some of the knowledge that separates those who know lots of mathematics from those who know only a little is straightforwardly empirical. A set theorist typically knows which axioms of set theory are generally accepted within the mathematical community, which theorems have been proved, which unsolved problems are generally regarded as important; and any algebraist knows that mathematicians have thoroughly developed a generalization of vector space theory in which the role that fields play in vector space theory is played by unitary rings. All these sorts of knowledge are empirical knowledge about the mathematical community. Besides such knowledge about the mathematical community, mathematicians typically have other empirical knowledge that non-mathematicians tend to lack; for example, typically there will be various complicated empirical claims about physical space which they know because they know them to follow from the empirical fact that physical space can be locally approximated as a Cartesian power of the real numbers. If one were to attempt a realistic account of all of the knowledge differences that separate a typical mathematician from a typical non-mathematician, I think that such differences in empirical knowledge would play a large role. So (contrary to what the title of this paper might suggest) a great deal of mathematical knowledge is the sort of straightforward empirical knowledge that no one could possibly regard as logical. The interesting question, however, concerns the mathematical knowledge that remains when this straightforward empirical knowledge is ignored. The deflationist claim that I have defended is that the only such knowledge there is is purely logical – even on a conception of logic according to which logic can make no existence claims. (It would not seem to me to be terribly interesting to say only that such knowledge was logical on a broader conception of logic like that of the logicists – a conception on which logic guarantees the existence of a realm of platonic entities. See footnote 3.)

The deflationist view is reminiscent of, but importantly different from, a position that has been called 'deductivism' or 'if-thenism.' Deductivism is usually characterized as the view that when someone asserts a typical mathematical statement (e.g., that there are infinitely many primes), what he or she really means is that this statement follows from a certain body of other mathematical statements.[37] (*Which* body of other

[37] See, for instance, ch. 3 of Michael Resnik, *Frege and the Philosophy of Mathematics*.

mathematical statements? The standard axioms of the field of mathematics in question, if the field has an accepted axiomatization; otherwise, some other body of claims implicit from the context. Deductivists tend to be a little vague about this, often entirely ignoring the situation where there is no generally accepted axiomatization.)

One of the major differences between this and deflationism is that deflationism, unlike deductivism, does not claim that mathematical claims mean anything other than what they appear to mean. Instead of saying that mathematical assertions don't mean what they appear to mean, the deflationist says that what they literally mean can't be known: the knowledge that underlies a mathematician's assertions is not what those assertions literally say. I'm afraid that many readers will still find this implausible, but it certainly seems to me less implausible than the claims about meaning made by the deductivist. (The implausibility of the deductivist position is especially evident in the case of a mathematical assertion A made in the absence of a generally accepted axiomatization. The deductivist must select some one body of other mathematical statements, and claim that what is meant in saying A is really that A follows from this other body of statements. But the deflationist, since he or she makes no claim about meaning, need not single out any one body of other mathematical statements as relevant: *no* bodies of other statements are relevant to what the assertion of A *means*, and *lots* of bodies of assertions are relevant to what the mathematician who asserted A *knows*, since a great many distinct pieces of knowledge of the interrelation of A with other mathematical claims may have been part of the motivation for asserting A.)

Indeed, the deflationist can easily handle a problem that is often thought to sink deductivism. Consider a mathematician asserting a claim which he or she knows not to follow from previously accepted axioms; he or she intends it as a new axiom. What does he or she mean when asserting this mathematical claim? If one takes deductivism entirely literally, then according to the deductivist, the mathematician must mean either

(a) that the new axiom follows from the old axioms

or

(b) that the new axiom follows from the system that consists of the old axioms plus the new axiom.

But the mathematician doesn't believe (a), and (b) is totally trivial. Since the whole point of the deductivist's claims about what the mathematician means is to make what he or she means directly reflect part of the knowledge that led to the assertion, both of these alternatives are intolerable.

For a deflationist, on the other hand, the situation where the mathematician introduces a new axiom poses no special problem. The reason is that the deflationist does not accept the programme of trying to represent the knowledge that leads the mathematician to make the assertion in the meaning of the assertion. The kind of knowledge that typically leads a mathematician to assert a new axiom is clear enough: it is knowledge that the axiom (in conjunction with previously accepted axioms) has certain desirable consequences and doesn't seem to have undesirable ones. In other words, it is knowledge of the logical interrelations of the proposed axiom with other mathematical claims, which is just the sort of knowledge that the deflationist wants to allow anyway. So the situation where someone asserts a mathematical claim because of the consequences he or she knows it to have is no more of a problem for a deflationist than the situation where he or she asserts it because he or she knows it to be a consequence of previously accepted claims. In general, I think it would be extremely surprising if careful attention to the sociology of mathematical practice turned up features of that practice that couldn't plausibly be handled along deflationist lines.

6

It is often supposed that one of the things that differentiates those who know a lot of mathematics from those who know only a little is that the former have considerable knowledge of a realm of platonic entities such as sets, numbers and tensors – entities that bear neither causal nor spatio-temporal relations to us or to anything we can observe. If this were correct, there would be a considerable problem in explaining how knowledge of such a realm could be attained. The strategy of this paper has been to point out various facts that aren't about such a platonic realm, facts which mathematicians typically know and non-mathematicians typically don't. These facts include empirical facts, like the facts mentioned early in section 5; and they include logical facts, like the facts about logical possibility that I have stressed in the bulk of the paper. I see no reason to believe that there is a further kind of fact – non-empirical and non-logical – that the mathematician also knows. Therefore, I see no reason to suppose that the mathematician has knowledge of the existence of mathematical entities or the truth of ordinary mathematical claims.

Appendix

In this appendix I will sketch an alternative to Kripke's model theory for modal logic, one which will give the intuitively correct results about which sentences involving the 'logically possible' operator are logically true. The model theory will be, like Kripke's, platonistic, for it will presuppose a large body of pure set theory. What a deflationist should say about the status of a platonistic model theory such as this is discussed in section 4 of the essay. (The purpose of the model theory is not to confer intelligibility on the modal operator. In my view, logic stands on its own; it doesn't need model theory for its intelligibility. Indeed, it is hard to see how a logic could get its intelligibility from the model theory for it, since one would need the logic in understanding (e.g., in being able to reason from) the model-theoretic assertions.)

In the model theory for first order logic, we say that a sentence is *logically true* if and only if it is *true in all models*. Here, a *model* consists of a set of objects (the entities that *exist in the model*) plus a stipulation as to which things if any the predicates are *true of* in the model, which things if any the names *denote* in the model and so forth. We also need the notion of an *assignment function for a model*: it is a partial function that assigns entities that exist in the model to all, some or none of the variables of the language.[38] Given a model M and an assignment function

[38] It is standard in first order logic to restrict consideration to models in which at least one object exists and in which all names denote. When this restriction is made, assignment functions are taken to be total functions, that is, they assign something to every variable. I have tacitly lifted this restriction in the body of the appendix, since it would be more glaringly anomalous in modal logic than it is in first order logic. (There are two ways of lifting the restriction that seem about equally reasonable, those of Scott and of Burge – see fn. 13. Strictly speaking, the definition of 'model' given in the appendix is applicable only to Burge's system, but nothing of substance would be altered if we complicated the definition slightly so as to apply to Scott's. In particular, the definition of 'model' for a modal logic based on Scott's free logic would be the same as for the non-modal Scott free logic.)

s for M, it is possible to recursively define what it is for a formula of the language to be *true in M relative to s*. We then define a formula to be *true in M* if and only if it is true in M relative to s, for *every* assignment function s for M; and we define a formula to be *logically true* if and only if it is true in M for every model M.

How should we generalize this to modal logic? Kripke's approach is to keep the idea that logical truth is truth in all models, but to redefine the notion of model: in the case of the system S5 (which is the one of interest for present purposes), a model is to be a non-empty set of possible worlds, one of which is designated as actual; each possible world is determined by a set of objects that exist in that possible world, plus a stipulation as to what things in that world the predicates are true of in that world in that model, plus similar stipulations for names and other primitive vocabulary of the language. A sentence of the form '◇A' will be true in a model just in case A is true in at least one possible world in the model. For '◇A' to be logically true, it must be true in all models, *and hence in particular it must be true in all models in which there is only one possible world* (i.e. in which there are no possible worlds other than the actual world of the model). It is clear that there is no way that this can happen unless A itself is true in all models, that is, unless A itself is logically true. That is the curious feature of Kripke's definition of logical truth for modal logic that I noted at the beginning of section 2.

I propose an alternative way of generalizing the definition of logical truth for sentences of first order logic to a definition appropriate to modal sentences. As on Kripke's approach, we are to retain the idea that logical truth is truth in all models. In addition, *we retain the definition of model used in first order logic*. (We do not introduce possible worlds; rather, a model will be in effect just the 'actual world portion' of a Kripke model.) The *only* difference between the proposed definition of logical truth for modal logic and the usual definition of logical truth for first order logic is that in recursively defining truth in a model we need an extra clause that will handle formulas beginning with '◇'.

Moreover, the rule for '◇' will be a lot like the rule for '∃' used in first order logic. In first order logic the rule for '∃' is as follows:

> '∃xB' is true in M relative to s if and only if B is true in M relative to s*, for some s* that assigns something to the variable x and that is just like s except in what it assigns to the variable x.

Note that in this rule we quantify over assignment functions, leaving the model fixed. I propose that in our rule for '◇' we quantify over models and assignment functions together:

'◇B' is true in M relative to s if and only if B is true in M* relative to s*, for some model M* and some assignment function s* for M*.

If B is a sentence (i.e., has no free variables), all reference to assignment functions can be proved irrelevant. That is, in that case the rule reduces to

'◇B' is true in M if and only if B is true in M* for some model M*.

It is clear that on this approach, unlike Kripke's, such sentences as

$\Diamond \exists x \exists y (x \neq y)$

and

$\Diamond \exists x$(x is an electron)

will come out true in all models, and hence logically true.

The approach that I have sketched for defining logical truth for modal sentences has its roots in chapter 5 of Carnap's *Meaning and Necessity*;[39] though much of what Quine found abhorrent about Carnap's ideas has been dropped. In the first place, the approach I have sketched does not rely in any way on the idea of meaning. As I remarked in the text, the basic modal logic is formulated in such a way that it does not reflect 'meaning relations among predicates' if such a notion be recognized. (Though if one does recognize such a notion, a derivative modal logic can be obtained which does reflect such relations.) In the second place, the treatment of free variables that I have given does not require the introduction of 'individual concepts', and it is thoroughly anti-essentialist in that no formula of the form '◇B' is true in a model with respect to one assignment function unless it is true in that model with respect to every other assignment function. (Again, however, the notion of logical

[39] An account even more similar than Carnap's to that given here is that of Nino Cocchiarella, 'On the primary and secondary semantics of logical necessity'. But Cocchiarella's method of dealing with variables and their interaction with modal operators seems to me unacceptable: for instance, it leads to Ramsey's bizarre conclusion that 'It is possible that there are at least $10^{10^{10}}$ objects' is logically false if the world happens to contain less than $10^{10^{10}}$ objects, but is logically true if the world happens to contain at least $10^{10^{10}}$ objects. (Cf. the last section of Ramsey's paper 'The foundations of mathematics'.) Despite this difference between Cocchiarella's account and the one I have offered, most of Cocchiarella's philosophical remarks about logical truth in modal logic apply to my account as well as to his own. (Some of the others who have advocated something in this general ballpark are Richard Montague (*Formal Philosophy*, ch. 1); Jaako Hintikka ('Standard vs. nonstandard logic') and Dana Scott ('On engendering an illusion of understanding'). Charles Parsons discusses a similar viewpoint in 'Quine on the philosophy of mathematics'.)

possibility can be used to introduce derivative notions of possibility which are essentialist in various ways.) A third respect in which my views modify Carnap's (though in this case, not a modification that Quine would favour) is that Carnap's idea was to regard modal concepts as derivative from semantic concepts. On my view it is, if anything, the other way around, as long as it is purely logical possibility that is in question.

Postscript

1 **Motivation**. The paper does not sufficiently emphasize that the idea of modal metalogic is appealing independently of anti-platonism. For more on this, see section 5 of the introduction to this volume.

2 **Substitutional Quantifiers**. In section 3 of this essay I used 'substitutional quantifiers' (viewed as devices of infinitary conjunction) for two purposes. But they weren't really needed.[1]

Mathematical and physical theories are standardly axiomatized with finitely many axioms and in addition finitely many axiom schemata each of which has infinitely many instances. The use of such axiom schemata is usually thought (rightly or wrongly) to make sense independently of any devices (such as substitutional quantifiers) which would enable us to encapsulate the schemata into single axioms. The idea is that even without reducing such schemata to single axioms, we can explain what it is to accept the schema: to accept the schema is simply to accept each of its instances. Let us assume here that this usual attitude is correct.[2]

The first use to which I put substitutional quantifiers in this essay was to express modally the idea that a given infinitely axiomatized theory in mathematics or physics was consistent. What is primarily of interest here (I would argue) is mathematical and physical theories expressed in a logic (such as first order logic) for which compactness holds. In this case, the consistency of the whole theory

[1] In the original version I made another use of substitutional quantifiers in section 4, but I have dropped that discussion from the current version, as explained in the text.

[2] There are two possible grounds for questioning it. One is that the idea of accepting infinitely many things makes no sense, unless there is a finite bunch of principles that we accept from which they obviously follow. Another is that even if we do suppose that it makes sense, still accepting the instances is too weak an account of what it is to accept the schema: accepting the schema is like accepting the conjunction of all the instances, and we could accept each instance without accepting the infinitary conjunction. I sympathize with both these grounds, and as a result think that the use of substitutional quantifiers as a device of infinitary conjunction is ultimately needed in metalogic. But my point here will be that the need has nothing to do with the issues about modality and consequence *per se*.

is the same as the consistency of each of its finite conjunctions; so we can explain what it is to believe that T is consistent by saying that we believe each sentence $\Diamond(T_1\& \ldots \&T_n)$, where T_i s are axioms. The 'infinitariness' involved in accepting the consistency of T seems no more problematic than the 'infinitariness' of accepting T itself. If we have an infinitely axiomatized theory that we believe is logically true, then expressing our belief is also unproblematic: we believe each claim of form $\Box T_i$. (These beliefs of course entail $\Box(T_1\& \ldots \&T_n)$, for each finite conjunction).

If we want to deny the consistency of an infinitely axiomatized theory, though, and if we don't know where the inconsistency lies, then we have a bit of a problem. The inconsistency of T is of course the same as the logical truth of the denial of T; but we may not know how to express the denial of T without conjoining the sentences of T, and consequently we may not be able to deny the consistency of T in this case. Here substitutional quantifiers (or some more or less equivalent device of infinite conjunction, such as a disquotational truth predicate – see section 4 of essay 7) are needed. Note, though, that the reason they are needed doesn't have anything to do with the modal representation of consistency *per se*, it has to do rather with the representation of *negation* for infinitely axiomatized theories. We would have the same problem if we wanted to express the belief that an infinitely axiomatized theory was false if we didn't know which part of it was false.

At any rate, in this essay the issue of denying the consistency of an infinitely axiomatizable theory never arises.

The second use to which I put substitutional quantifiers was in formulating the strong consistency or conservativeness of a mathematical theory. But again we don't really need substitutional quantifiers: to accept the conservativeness of M is simply to accept each instance of the schema

(C) If $\Diamond B$ then $\Diamond(B^* \& M_1 \& \ldots \& M_n)$,

where B is any sentence, B^* is the result of restricting B to non-mathematical entities, and $M_1, \ldots, M_n$ are axioms of M. To be sure, this only directly expresses the conservativeness of M with respect to *finitely axiomatized* nominalistic theories B;[3] but the conservativeness with respect to infinitely axiomatized nominalistic theories follows. Proof: suppose T is infinitely axiomatized and consistent, i.e., $\Diamond(T_1 \& \ldots \& T_m)$ whenever $T_1, \ldots, T_m$ are any of its axioms. If M is conservative with respect to finitely axiomatized theories, then (since the starring operator distributes over conjunction) $\Diamond(T_1^* \& \ldots \& T_m^* \& M_1 \& \ldots \& M_n)$ for any $T_1, \ldots, T_m, M_1, \ldots, M_n$. But this is just what it should mean to say that $T^* \& M$ is consistent.

The upshot of this is that while I believe in the use of substitutional quantification for certain purposes (viz., the purposes that others would use disquotational truth for), they aren't really needed in the context of this essay, and it would have been tactically advantageous simply to formulate conservativeness by schema (C).

[3] (C) holds even for non-nominalistic B: the effect of the * operator is to reinterpret such a B in an unintended way as nominalistic.

3 **Modal soundness and conservativeness.** I should have pointed out that my modal soundness principle MS is really the platonistic formulation of another kind of conservativeness claim, and that MS# is the deflationist's version of the same claim. What MS and MS# amount to is that typical applications of proof theory don't yield conclusions you wouldn't get otherwise, just as what I've been calling conservativeness says that other typical applications of mathematics don't yield conclusions you wouldn't get otherwise. It is illuminating to write the deflationist's version of both kinds of conservativeness together, in a common format. MS# generalizes from NBG to mathematical theories more generally as the schema

(MS#) If $\Box(M_1$ & . . . & $M_n \supset$ there is a proof of '-A' in F), then $-\Diamond$A,

while (C) contraposed yields

(C) If $\Box(M_1$ & . . . & $M_n \supset$ -A*), then $-\Diamond$A.

4 **Strengthening MTP# and MC#; the modal version of the Kreisel squeezing argument; and modal analogues of provability**. The modal derivation of MTP# (from the Conditional Possibility Principle) proceeds by the derivation of a stronger claim:

(MTP*) $\Diamond$(NBG & there is a model of 'A') $\supset \Diamond$A.

(It's stronger given $\Diamond$NBG, anyway.) From this and classical completeness, we get a strengthened modal completeness principle:

(MC*) $\Diamond$(NBG & there is no F-derivation of '-A') $\supset \Diamond$A,

where F is a typical formal system for quantificational logic. Analogous strengthenings of MS and ME, by contrast, are false, as footnote 24 shows. (In the original version of this essay I asserted an analogue of ME* *when NBG was replaced by a strengthened mathematics that included an ω-rule*; but here I am dropping that strengthening.)

In this essay I do not make anything of the fact that we can derive MTP* rather than merely MTP#, but it is important. Consider the Kreisel analysis (introduction, section 5) of the philosophical significance of the completeness theorem for first order logic. Kreisel argues in effect that by taking 'it is logically consistent that' as a primitive operator (here symbolized as '$\Diamond$') governed by the principles MTP and MS, the role of the completeness theorem is to enable us to prove the biconditionals

$\Diamond$A $\equiv$ there is as model of 'A'

and

$\Diamond$A $\equiv$ there is no F-derivation of '-A'.

We *prove* these, without regarding either as *defining* consistency. This is of course a platonistic analysis, since an anti-platonist cannot accept those biconditionals, but it would be nice if the anti-platonist could say something analogous. The analogue will of course involve using MS# instead of MS; and we get an especially clean analogue if we use MTP* (rather than MTP#) instead of MTP. The analogue goes like this: MS# gives as a *necessary* condition for ◇A that ◇(NBG & there is no F-derivation of '-A'); MTP* gives as a *sufficient* condition for ◇A that ◇(NBG & there is a model of 'A'). But the completeness theorem in the form the anti-platonist accepts it – that is, □{NBG ⊃ (there is no F-derivation of '-A' ⊃ there is a model of A)} – shows that the necessary condition entails the sufficient condition: there is no room between them, so each is both necessary and sufficient. That is,

(i) ◇A ≡ ◇(NBG & there is a model of 'A')

and

(ii) ◇A ≡ ◇(NBG & there is no F-derivation of '-A').

Or equivalently,

(iii) □A ≡ □(NBG ⊃ 'A' holds in all models)

and

(iv) □A ≡ □(NBG ⊃ there is an F-derivation of 'A').

From a platonist point of view, it should be no surprise that these hold. Consider (iv): the platonist analysis using MTP and MS yields

□A ≡ there is an F-derivation of 'A';

but a platonist will also accept

there is an F-derivation of A ≡ □(NBG ⊃ there is an F-derivation of 'A'),

since 'there is an F-derivation of *x*' is a Σ_1 formula, and since NBG is ω-consistent. (1-consistency is actually all that's needed.) The two indented conditions give us (iv). Of course, the derivation of (iv) in the previous paragraph was independent of this platonist derivation of it: it relied only on the Conditional Possibility Principle (to get MTP*) and MS#. (The use of MS# is the anti-platonist analogue of the assumption of 1-consistency.)

Despite the fact that the operators ◇ and □ in (i)–(iv) occur on the right hand side of the biconditionals as well as on the left, I think that (i)–(iv) should allay any doubts about the clarity of those operators. (At least, they should allay doubts about the clarity of the operators as applied to non-modal sentences, which is where I primarily want to apply them.)

Incidentally, since 'there is an S-derivation of *x*' is Σ_1 for any formal system S (not just for formal systems of quantification theory), then from a platonist point of view we have in general that □(NBG ⊃ there is an S-derivation of '-A') if and only if there is an S-derivation of 'A'. (The same holds if 'NBG' is weakened to 'platonistic proof theory'.) In footnote 24 I raised two difficulties for a modal definition of S-derivability. The kind of definition I was considering there was

◇(Proof theory & there is an S-derivation of 'A'), and the second difficulty was that (because of Gödel's theorem) this has to be extensionally inadequate if the proof theory is formulated in first order logic alone. We now see that in a sense we could have solved the difficulty by a different kind of modal 'definition' of S-derivability, namely

$\Box$(Proof theory $\supset$ there is an S-derivation of 'A').

It should be noted, though, that this has no plausibility whatever as an account of the ordinary notion of derivability: it is a modal *surrogate* of derivability, not a modal *analysis*. If one wants a modal analysis, one still has to introduce richer logical structure to go with the modality, as discussed in footnote 24.

5 On the strong modal logic employed. In the appendix to this essay I sketch a platonist model theory for a quite strong modal logic: it consists in tacking onto S5 *all truths of form '$\Diamond$A', where A is non-modal.* (More exactly, one tacks all such truths onto a certain anti-essentialist version of quantified S5.) However, no positive use is made of anything but an axiomatized fragment of this.

I do need a strong possibility axiom, such as $\Diamond$(NBG). (It might be better to try to find more obvious possibility principles from which this could be shown to follow, but I have not felt the need to pursue this: $\Diamond$(NBG) seems obvious enough.) I also need two other principles: the Conditional Possibility Principle of section 2 and, ultimately, MS#. (These are consequences of the strong logic with all true possibility statements in it, but they need separate inclusion with an axiomatized fragment.)

I did make a negative use of the strong logic with all true possibility statements in it: I argued that the modal soundness claim employed by the platonist, namely MS, isn't a modal consequence of standard mathematics even in that strong logic. But of course here the only role of the strong logic is to serve as an upper bound on the logic that might reasonably be employed.[4]

It is worth pointing out that to define logical truth in such a way that it obeys a strong modal logic does not commit one to any strong claims about logical *knowability*. And in this essay I make a sharp distinction between logical truth and logical knowability: indeed, most of section 2 is devoted to considering the view that even if claims like $\Diamond$(NBG) are logically true, they cannot be logically known. Of course, I considered the view only to reject it. But my argument against the view was certainly *not* an argument that the distinction between logical truth and logical knowability collapses.

[4] The only other role that I gave to the strong logic was in my extension of MTP and ME from non-modal formulas to modal formulas: it is of course hardly surprising that one would need the strong modal logic there, since MTP and ME, as applied to modal formulas, are modal principles about truth of modal formulas in a model, and the definition of model and of truth in a model for modal formulas were designed with the strong modal logic in mind. Presumably if one wants to isolate a weaker modal logic one could find a notion of model and of truth in a model that is appropriate to that weaker logic, and given such a construal of truth in a model, MTP (and perhaps ME) could be shown in the weaker logic to hold for modal formulas as well as for non-modal ones. But I see little reason to pursue this: indeed, I'm not sure that MTP and ME are of much importance in their application to modal formulas.

6 Logical Knowledge. My claim that mathematical knowledge is just logical knowledge (insofar as it isn't just knowledge about the mathematical community or knowledge-how instead of knowledge-that and so forth) may suggest to some the idea that mathematical knowledge is indefeasible. This would be false (as the discovery of the inconsistency of Cantorian set theory illustrates dramatically), and was not intended: for I take logical knowledge to be defeasible too. In particular, the paper argues that claims of consistency are to be construed as logical, and they are clearly defeasible. (In the essay, I attempted to remove any suggestion of indefeasibility, by sometimes shifting from claims of what we can logically *know* to claims of what we can have logical *reason to believe*.) Also, the claim that knowledge or rational belief is logical was not meant to preclude its having a somewhat inductive character: indeed, I argued that knowledge of consistency of certain theories is at least partly based on the idea that if the theories were inconsistent we would probably have discovered it by now, but that this didn't prevent the knowledge from being logical. What I primarily meant to be saying, in calling knowledge of the consistency of mathematics logical, was that this knowledge did not have to be based on knowledge of that mathematics itself (or some stronger mathematics; or some theory about other entities as epistemologically problematic as mathematical entities, such as possible worlds).