Is Arithmetic Consistent?

GRAHAM PRIEST

Introduction

Let L be the language of first order arithmetic; and let N be the set of sentences of L true in the standard interpretation. It is well-known that N has many (in fact, absolutely infinitely many) models other than the standard interpretation. All of these models extend the standard interpretation in a certain sense, and have an interesting common structure.[1]

What is less well-known is that research into the model theory of paraconsistent logics has established that N has many (countably infinitely many) models of which the standard interpretation is itself an extension, in a certain sense, and which also have an interesting common structure. The models are finite, but also verify *all* the truths of the standard model. It is not my aim to investigate the models of N here. Rather, I want to explore the consequences of their existence for some issues in the philosophy of mathematics.[2]

1. Inconsistent Arithmetics

Let me start by spelling out the main technical result I will be appealing to, which goes as follows. For each natural number, n, there is a set of sentences of L, N_n, with the following properties:

i) $N_n \supseteq N$, and so N_n is complete (i.e., for every sentence φ, N_n contains either φ or $\neg\varphi$).
ii) N_n is a theory in the paraconsistent logic LP.
iii) N_n is inconsistent.
iv) If φ is a (negated) equation concerning only numbers $< n$ then $\varphi \in N_n$ iff $\varphi \in N$. (Hence if $n>0$ N_n is non-trivial.)
v) N_n is decidable (and so axiomatisable).
vi) N_n is representable in N_n, and hence L contains a truth predicate for N_n.
vii) If B is the proof predicate for N_n then every instance of the scheme $B(<\varphi>) \rightarrow \varphi$ is in N_n.

[1] See, e.g., Boolos and Jeffrey (1974), Ch. 17.

[2] The first person to note that the models have important *philosophical* implications was van Bendegem (1991). This paper draws somewhat different conclusions from his, but is much indebted to it.

Mind, Vol. 103 . 411 . July 1994

viii) If φ is any non-theorem of N_n, $\neg B(<\varphi>)$ is provable in N_n. Hence the non-triviality of N_n can be established in N_n (by finitary means).

ix) The "Gödel sentence" for N_n is (provable) in N_n, as is its negation.

A sketch of this result can be found in Appendix 1. For those who are curious, but do not want to work through the technical details, let me indicate how the trick is turned. N_n is the set of sentences true in the LP model constructed as follows. The domain of the model is just $\{m; m \leq n\}$. Arithmetically, all the numbers less than n behave in the standard way, but n has any atomic or negated atomic property iff some number $\geq n$ has it.[3] Of course, n is an inconsistent object (in fact, the least inconsistent object) in the interpretation. In particular, in the model, **n=n+1** is true even though it is also false (I write the numeral for i as **i**). (i) is established by appealing to a result known as the Collapsing Lemma. (ii), (iii) and (iv) are immediate. (v) follows simply from the finitude of the interpretation. (vi) and (vii) and (viii) follow from (v) and the fact that all decidable sets are definable in N, and so in N_n. (ix) requires a little more (but straightforward) computation.

2. *The standard interpretation*

Let us, henceforth, fix n as some incredibly large number, say a number larger than the number of combinations of fundamental particles in the cosmos, larger than any number that could be sensibly specified in a lifetime, so large that it has no physical meaning or psychological reality. Let us call N_n, M. The central question I want to ask in this essay concerns the theories N and M, and is simply: which of these is correct, i.e., which of these is the complete set of truths of arithmetic? Of course, N is the complete set of truths in one interpretation and M is the complete set of truths in another. The question, therefore, is: what is the correct interpretation of the language L?

The question might appear silly, since the answer is so obvious. Clearly, the standard interpretation of L is the correct one. Hence, it is N that is true. The first thing I want to do is to show that things are not this straightforward. The argument is, in fact, a well-known one, but with a small twist. Observe, first, that the natural way to understand talk of the standard interpretation is platonistically: an interpretation is an abstract structure, to parts of which the terms of the language refer, and over which quantifiers range. If the notion is understood in some other way (which may, indeed, be possible), it is not at all obvious what this is, and so not obvious that the matter is simply settled.

Next, forget the inconsistent interpretations of L for a moment, and just recall the multitude of consistent interpretations. Given that there is such a multitude, what is it that determines that our language has any particular one of them as its

[3] The specification of N_n given here is parasitic on an understanding of the standard interpretation. This feature is inessential, however. There are independent (though pedagogically more complex) specifications.

unique interpretation? It is certainly nothing that we can assent to in L, for all the interpretations are elementarily equivalent (make the same things true).[4] Arguably, there is nothing.

This argument has, of course, been used by Hilary Putnam (1983), who applies it much more generally. I have never been persuaded by the generalisation of the argument to the interpretation of physicalistic language. In this case, it seems only too obvious that there are things that select the correct interpretation other than the set of statements endorsed. For example, there are independent factors that fix the referents of the terms involved. Exactly what these are, one might argue about; however, reference-fixing is not done by magic, and the most plausible account of how reference is fixed seems likely to involve appeal to some causal relation between the referent and the speaker.[5] However good this reply is for physical language, it is obviously not available for the language of arithmetic: numbers are not causal agents.

As is well-known, Gödel argued that we have a mental faculty of direct intuition for numbers. I do not think that this metaphor ultimately has any cash value (Priest 1987, p. 189), but let us grant it here. Could *it* be what determines reference? I do not think so. Any thesis to this effect would seem to fall foul of the Private Language Argument in one form or another. If the fixing of reference is performed solely by subjective mental acts, then there is nothing to prevent each of us fixing reference in a quite different way, which is another way of saying that *qua* public language, reference is not fixed at all. L is, after all, a public and shared language; the criteria for reference-fixing must, therefore, equally be public.

Perhaps there are other candidates for reference-fixing; however, I know of none worth discussing. Let us, then, recall that in the present situation there are possible interpretations *other than* the consistent ones. Does the existence of these materially alter the situation? No; if anything, it strengthens the argument. Putnam's argument appeals, essentially, to the Löwenheim-Skolem Theorem. The Collapsing Lemma that is used to construct the finite interpretations of arithmetic is the ultimate downwards Löwenheim-Skolem Theorem: arithmetic has a model of *every* cardinality. The argument from multiplicity to indeterminacy is therefore reinforced.

It is true that N is not a *complete* description of the finite interpretation, in the sense that there are things true in the interpretation that are not in N. But this does not undercut the argument: it merely reinforces it. If this is not clear, just compare

[4] Nor does it help to consider extensions of L, e.g., with set theoretic notions. There is just the same multitude of models of, e.g., classical ZF with non-standard natural numbers. This is no longer true if we interpret the set theoretic machinery as second order. Interestingly, however, it is till true in the inconsistent case. The proof of this is a straightforward extension of that given in Appendix 1.

[5] Putnam, of course, replies to this move that it is "just more theory". This reply seems to me to fail: the point is that the causal relationship is there to determine; that we might describe it in ways that themselves may be reinterpreted is beside the point. See, e.g. Lewis (1984), p. 225.

it with the classical situation: N is our best description for characterising the interpretation of L; if that can't do it then incomplete theories, such as the set of theorems of first order Peano Arithmetic, P, certainly can't. From the perspective of a finite interpretation, N is incomplete, just as P is, classically; but this fact has no relevance. Of course, we might argue that we have reason to suppose that N *is* complete in the pertinent sense. But this is a different argument, which we will come to in due course. The point of the present section is merely to argue that there *is* a substantial issue here; and the case for this seems made.

3. Practice, application and consistency

If appeal to abstract structures cannot settle the matter of the correct interpretation for L, what else can? Another candidate is our rule-following practices of counting, adding, etc. Perhaps this can settle the matter.[6] Let us consider two people, α and β, whose practices support N and M respectively. How do the practices of α and β differ? Both would count in the same way: 0,1,...*n*, *n*+1...; both would add and multiply using the same rules; both would say, for any *m*, that *m* is different from 0, 1,...*m*-1, *m*+1,... But β would also say for any $m \geq n$, that *m* is indentical to *n*, *n*+1,..., whilst α would not.

Surely, you and I are like α, not β? Although one is intuitively inclined to say "yes" to this, the answer is not so obvious once one starts to think about it. As Wittgenstein demonstrated, any determinacy there is in the notion of rule-following is to be grounded in the fact that we have dispositions to proceed in a socially universal (or at least, pretty common) way.[7] Both α and β proceed in the same way for all *actual* situations. The divergence between them could appear only in situations that transcend anything humanly possible. What makes one think that in such situations one would behave like α, rather than β? Our knowledge of how we would proceed in hypothetical situations is notoriously unreliable. Even worse, it is not even clear that there is any fact of the matter here. What sense is there in the notion of a human disposition to act in a situation that is humanly impossible? (Kripke 1982, p. 26f.) Considerations such as this are apt to undercut one's (or at least my) initial inclination to say that I am like α, not β.

Having put aside both the transcendental and human practice, there doesn't seem much left to guide a decision between N and M except the natures of N and M themselves. What scope is there here? The disagreement between exponents of N and M is not of the usual kind. *Anything* (at least anything in L) that the one says, the other will agree with. All that the one can do is to mark disagreement with what we might call the surplus content of the other's views. For example, if M is true, then there is a largest number (as may easily be checked). The one will

[6] A similar suggestion is made by Wright (1985, p. 130), in the context of the classical debate surrounding the Löwenheim-Skolem Theorem.

[7] See, *Philosophical Investigations*, especially §§201-242. See also Kripke (1982, Ch. 3).

deny this. But this hardly settles the matter: the other, who endorses M, will, of course, agree that there is no largest number as well. If the one has a move at this point, it would seem to be that the other is inconsistent. But what, in this context, is wrong with this? A standard argument is to the effect that if arithmetic were inconsistent, we would lose all control over it, since everything would follow. It is precisely this view to which the existence of M gives the lie. And most of the arguments for consistency—that I can think of, anyway—beg the question in a similar way.

One non-question-begging argument appeals to the application of arithmetic. If arithmetic were inconsistent then, surely, in applying it to build bridges, for example, we could expect them to fall down—which they don't.[8] Leaving aside the fact that bridges *do* fall down sometimes, the argument carries little weight here. The portion of arithmetic that is applied in engineering and other parts of human life is largely computational; and not just computational: the computations involve magnitudes that are physically meaningful. Now N and M *agree* on all computational mathematics—at least up to n, and what happens beyond this is, by choice of n, of no physical import.

4. Petersen's Argument

I don't pretend to have aired all the arguments in favour of N over M. I am sure there are others. But the discussion so far will at least, I hope, have shown that one ought to be open-minded enough to see what the other side of the case is like. The first thing to note here is that there is a very simple reason to favour the unorthodox view. Using an argument due to Uwe Petersen (personal communication), we can *prove* that there is a number, x such that $x=x+1$![9] I will give an informal version of the argument here; a formal version can be found in Appendix 2.

Let π be an abbreviation for the following description:

> the least number such that this description refers to it (or 0 if it fails to refer) + 1

Note that "π" does refer to a number. For if "π" failed to refer, it would refer to 1. Hence we have:

$$\pi = (\text{the least } x \text{ such that "}\pi\text{" refers to } x) + 1 \qquad (*)$$

Now, clearly, "π" refers to π, and since reference is unique π is the least thing that "π" refers to. Hence, by (*), $\pi=\pi+1$, and so:

$$\exists x x=x+1 \qquad (**)$$

What this argument shows is that, as a description of the arithmetic facts, N is incomplete; but M, which contains (**), gets it right. The argument does not tell

[8] This objection to inconsistency is put by Turing to Wittgenstein. See Diamond (1976, p. 211).

[9] Petersen obtained it by analysing the formulation of Berry's paradox in Priest (1983). (See also Priest (1987, 1.8).) I am sure that he would not approve of my use of it, however.

us *what* the x in question is, in any particularly illuminating way. *A fortiori*, it does not tell us that it is so large as to have no psychological or physical significance in the appropriate sense (though the fact that we find it impossible to produce a candidate for x suggests at least the former). But as an argument against N, it is decisive.

It can also be used to answer another objection against M. We fixed n as a certain number but, for all we have said so far, any one of an infinite number of numbers could play this role. If it is M that is true, why is it not one of the other theories N_m for such an m? The argument points to the arbitrariness of the chosen n. All that is required to answer it, therefore, is to choose n in a non-arbitrary fashion. This can now be done simply. The argument demonstrates that there is a number equal to its own successor. Let n be the least such number. This is the *obvious* bound for the consistent behaviour of numbers.

5. Truth and decidability

The orthodox-minded will, naturally, be suspicious of Petersen's argument. It is, of course, closely related to numerous paradoxes, and so it is natural to suppose that some standard solution to such paradoxes will cope with it. I have argued against such solutions at length elsewhere,[10] and to take up the issue here seems pointless, as well as taking us away from the main point of this paper. So let us leave these matters there.

Even if this argument be rejected, there is another—or better, family of others—which, though perhaps less than conclusive, is still very weighty. The main idea is simple. The Limitative Theorems of classical metamathematics[11] are usually regarded as, at best, disappointing, at worst posing nasty philosophical problems. With the exception of the Löwenheim-Skolem Theorem—which has already reared its head—paraconsistent arithmetic is free from all these Theorems, and so problems. The fact that a theory solves problems that beset its rivals is well recognised as speaking strongly in its favour.[12] What follows elaborates on this.

Let us start with Church's Theorem. The hope that we might have a decision procedure to solve mathematical problems goes back, at least, to Leibniz. The most ardent esperant this century was, of course, Hilbert (at one time). Hilbert hoped that by formalising mathematics, and in particular arithmetic, we would be able to establish a decision procedure for it. Church's Theorem showed this to be impossible. But by clause (v) of § 1, if M is the correct arithmetic, there exists just such a decision procedure (and a very simple—though exponential—one at that).

[10] See especially, Priest (1987, Ch. 1).

[11] For a survey, see Fraenkel *et al* (1973, p. 310ff).

[12] The point has been argued by Kuhn, Lakatos and, perhaps most comprehensively, by Laudan (1977).

This is not, strictly speaking, an argument for M, but it certainly makes M enticing.

Let us move on to Tarski's Theorem: classical arithmetic cannot contain its own truth predicate. Perhaps there is nothing problematic about this if we just regard it as establishing that arithmetic truth is a notion stronger than we might have hoped. However, there is more to it than that. For the proof of Tarski's Theorem demonstrates that nothing satisfying the necessary conditions for a truth predicate (specifically, Tarski's Convention T) can be incorporated into arithmetic. In other words, a *single* theory of truth and number is beyond our grasp. This *does* seem puzzling. Why should these two notions be immiscible; and how can they be immiscible since we mix them? The puzzle is resolved by M; by condition (vi) of §1, the language of M contains its own truth predicate.

Unsurprisingly, the issue of the solution to the semantic paradoxes, and especially the Liar Paradox, is raised here. The fact that we cannot have a unified account of number and truth means that a Tarskian "metalinguistic" solution to the paradoxes must ultimately be endorsed. This is highly problematic, as many have noted. By contrast, M provides a clean and simple solution. Since I have discussed many of the issues involved elsewhere (Priest 1987, Ch. 1), I will not discuss them further here.

6. Gödel's Theorems

Let us now turn to Gödel's incompleteness Theorems. These, of course, have been held to have all kinds of problematic or unpalatable consequences. Notably, then, Gödel's Theorems all fail for M. There is no space here to review *all* the damaging implications that the Theorems have been supposed to have.[13] So I will just review a few of the more notable ones.

The simplest form of the first Theorem is that arithmetic is not axiomatic. (M is, since decidability implies axiomatisability.) Since axiomatisability has been the methodological cornerstone of mathematics since Euclid, this result came as something of a blow. The hard problem posed by the result is not methodological, however, but cognitive. We appear to obtain our grasp of arithmetic by learning a set of basic and effective procedures for counting, adding, etc.; in other words, by knowledge encoded in a decidable set of axioms. If this is right, then arithmetic truth would seem to be just what is determined by these procedures. It must therefore be axiomatic. If it is not, the situation is very puzzling. The only real alternative seems to be platonism, together with the possession of some kind of sixth sense, "mathematical intuition". I have argued against this elsewhere.[14] All

[13] And I certainly do not agree with all of them, for example, that the Theorems refute a mechanist philosophy of mind. See Priest (1994).

[14] See Priest (1987, 10.4.)

that we need note here is that we have already seen in §2 that the Löwenheim-Skolem Theorem gives the lie to the claim that platonism can provide an account of how we interpret (grasp the meaning) of arithmetic language. Hence, this alternative fails.[15]

A stronger form of the first Theorem is to the effect that for any axiomatic arithmetic of a certain kind, we can actually produce a statement of arithmetic (the Gödel sentence) that is not provable in the theory, yet which we can prove to be true. How is it possible that our powers of proof can outrun any axiomatic system? (The problem is similar to the first, except that it concerns proof, rather than truth/meaning.) If M is correct, this problem, too, is avoided because of clause (ix) of §1.

I argued (1987, Ch.3) that the only reasonable conclusion that can be drawn from this form of the Theorem is that our proof procedures are inconsistent. This argument is confirmed if M is correct. By (ix) of §1, the Gödel sentence of M is provable in M, as is its negation. The theory is, therefore, inconsistent. We knew that anyway, of course; but this is just the inconsistency one should expect if the analysis cited is correct.

Let us now turn to the second Theorem. This is to the effect that the sentence that canonically asserts the consistency of any axiomatic arithmetic of a certain kind cannot be proved in the system. Of course, M is inconsistent, so as far as M goes, the question of a consistency proof does not arise. However, the parallel question of non-triviality *does* arise. M is non-trivial (i.e., is not the totality of all sentences of L) and this can be proved in M, as part (viii) of §1 tells us.[16]

The unprovability of the consistency (non-triviality) of arithmetic in arithmetic was certainly a negative result for many, in that it killed off Hilbert's Programme. Whether or not the present situation is a plus for M is therefore connected with the question of the importance of the viability of that Programme. Before I discuss this however, there is a related fact that is a definite plus for M. A standard way of proving the second Theorem is via Löb's Theorem. Let B be the proof predicate for a theory. Then if

$$B(<\varphi>)\rightarrow\varphi \qquad (***)$$

is provable, φ is provable.[17] It follows that if the theory is non-trivial, not all instances of scheme (***) are provable in the theory, though they are true in the standard model. This has struck many as odd. How is it that some truths as innocuous as those that (***) expresses *must* fail to be provable? The situation is rectified by M. As (vii) of §1 assures us, every instance of (***) is provable in M.

[15] This may not refute platonism itself, but it certainly makes it vulnerable to an application of Ockham's Razor.

[16] This is, in fact, one of the earliest results in investigations of finite models of arithmetic. See Meyer (1976).

[17] See Boolos and Jeffrey (1974, Ch. 16).

7. Hilbert's Programme

Let us, finally, turn to Hilbert's Programme. The main ideas behind this were as follows.[18] The finitary part of mathematics (roughly, the computational part) has a perfectly good procedural interpretation and needs no appeal to platonism to make sense of it. Other arithmetic statements are to be considered as having only an instrumental meaning. They are acceptable to the extent that, but only to the extent that, they simplify and unify our operations with finitary sentences. A guarantee of acceptability is to be provided by: (a) an axiomatisation of arithmetic (or more generally, mathematics), together with: (b) a proof of the fact that the non-finitary axioms are conservative with respect to the finitary statements, that is, that they cannot be used to prove non-finitary statements that are not directly provable. In the context within which Hilbert was working, this is equivalent to finding a consistency proof for the theory. And if the consistency proof is to have any force it can not use the methods it is supposed to be justifying. Hence a proof must use only finitary methods.

Gödel's Theorems were thought to have killed off both (a) and (b). M puts the whole situation in a quite different light. First, M *can* be axiomatised. Secondly, as (viii) of §1 tells us, *any* non-theorem of M can be shown in M to be a non-theorem. In particular, any untrue (in the interpretation for M) finitary statement can be shown not to be provable. Moreover, since M is decidable, the methods used are strictly finitary.

So is Hilbert's Programme vindicated? Maybe, maybe not. First, some true finitary statements in M, and in particular, some equations, turn out to be inconsistent (have true negations). Hilbert might not have been too happy about this, though if M *is* true arithmetic, this unhappiness can legitimately be set aside. More worrying is the fact that a non-triviality proof is not necessarily to be accepted at face value. Since the system is inconsistent, the fact that $\neg B(<\varphi>)$ is provable does not prevent φ from being provable. But third, and conclusively, the while point of the exercise was to justify non-finitary reasoning, i.e., reasoning that goes beyond the computational. But M is decidable. Hence, if M is correct, *no* correct reasoning goes beyond the finitary. From the perspective of M, Hilbert's Programme is not so much realised as rendered redundant!

However, we are still left with the thought that finitary arithmetic, i.e., now, all of it, has a perfectly good procedural interpretation. There is therefore a perfectly good account of how it is we grasp the meaning of arithmetical statements and establish them as true. In many ways, this just summarises the central benefits of M, as I have tried to bring them out in the preceding sections. In a nutshell, this argument in favour of M is that it removes the mystification from mathematics. Or, at least, arithmetic. To what extent the considerations of this paper extend to the rest of mathematics is a topic for another occasion.

[18] See Hilbert (1925).

Appendix 1

In this appendix I will give the proofs of the facts cited in §1. This can be done reasonably succinctly since many of the key results are already in the literature.

Let $\mathfrak{A}=\langle D,I\rangle$ be any first order interpretation. Let ~ be any equivalence relation on D that is also a congruence relation with respect to the functions involved. Define the LP interpretation,[19] $\mathfrak{A}^{\sim}=\langle D^{\sim},I^{\sim}\rangle$ to be called the *collapsed interpretation*, as follows. $D^{\sim}=\{[d];\ d\in D\}$, where [d] is the equivalence class of d under ~. For any constant, c, $I^{\sim}(c)=[I(c)]$. For any n-place function symbol, f, $I^{\sim}(f)([d_1],\ldots[d_n])=[I(f)(d_1,\ldots d_n)]$. For any *n*-place predicate, P (including, *nota bene*, identity), its positive/negative extensions in $\mathfrak{A}^{\sim}$ are respectively:

$\{\langle a_1,\ldots,a_n\rangle;\ \exists d_1\in a_1\ldots\exists d_n\in a_n\ \langle d_1,\ldots,d_n\rangle$ is in the positive/negative extension of P in $\mathfrak{A}\}$

Collapsing Lemma

If φ is true/false in $\mathfrak{A}$, φ is true/false in $\mathfrak{A}^{\sim}$.

Proof

The proof is by recursion over the structure of sentences. For details, see Priest (1991, §7). The proof given there is for a language without function symbols. The fact that ~ is a congruence relation is sufficient to ensure that for any term, t, $I^{\sim}(t)=[I(t)]$. The extension of the proof to languages with function symbols is then obvious. □

Now let $\mathfrak{A}$ be the standard model of arithmetic; let ~ be the equivalence relation that puts every number < *n* into its own equivalence class, and every number ‡ *n* into a single equivalence class. Let N_n be the set of sentences true in $\mathfrak{A}^{\sim}$.[20]

The Collapsing Lemma gives (i). (ii) is immediate. For (iii), consider the sentence **n=n+1**. This is true in the collapsed model, as may easily be checked; but its negation is also true by the Collapsing Lemma. Since we have done nothing to affect the denotations of numerals whose denotations are less than *n* (except lift their type), it is easy to check (iv).

It is also easy to check that for any formula, φ, the following are true in the collapsed model:

$$\exists x\varphi \leftrightarrow (\varphi(x/\mathbf{0})\vee\ldots\vee\varphi(x/\mathbf{n}))$$
$$\forall x\varphi \leftrightarrow (\varphi(x/\mathbf{0})\wedge\ldots\wedge\varphi(x/\mathbf{n}))$$

and hence, using a recursive procedure, that any formula, φ, is equivalent in the collapsed model to one without quantifiers, φ^*. This will give us (v):

Decidability

N_n is decidable.

[19] For a definition of LP interpretations, see Priest (1987, Ch. 5). Crucially, in such interpretations, all predicates have positive and negative extensions which may overlap; and sentences take one of the truth values {1}, {0}, {0,1}.

[20] This is not quite the way I defined N_n in §1. The interpretation given there is, essentially, that obtained (in a standard fashion) by choosing the least member of each equivalence class to do duty for it.

Proof

The following is a suitable decision procedure. To test φ, it is sufficient to test φ^*. We do this by assigning to each equation, $t_1=t_2$, the value v, where:

$$1 \in v \text{ iff } I(t_1) = I(t_2) \text{ or both are } \ddagger n$$

$$0 \in v \text{ iff } I(t_1) \neq I(t_2) \text{ or both are } \ddagger n$$

(This is the truth value of the equation in the collapsed model.) The formula φ^* is then tested by using LP truth tables. □

A similar proof of decidability is to be found in §3 of Meyer and Mortensen (1984), for a different class of finite models of arithmetic.

As is well-known, every decidable set (of formulae) is representable in arithmetic, i.e., if X is a decidable set (of formulae), there is a formula, φ, of one free variable such that:

$$\text{if } \alpha \in X \text{ then } \varphi(<\alpha>) \in N$$

$$\text{if } \alpha \notin X \text{ then } \neg\varphi(<\alpha>) \in N$$

(where $<\alpha>$ is the numeral of the code of α). By the Collapsing Lemma, every decidable set is represented by the same formula in every collapsed model. Since N_n is decidable, it is represented by some formula $B(x)$. This will give us (vi):

The T-Schema

Every instance of $B(<\alpha>) \leftrightarrow \alpha$ is true in the collapsed model.

Proof

Either α is in N_n or it is not. If it is, $B(<\alpha>)$ is true in the collapsed model, as therefore is the equivalence. If it is not, its negation is true in the collapsed model, as is $\neg B(<\alpha>)$; whence, again, the equivalence is true. □

Warning: note that for an arbitrary B that defines N_n in the collapsed model, there is no reason to suppose that it satisfies the stronger identity condition:

α and $B(<\alpha>)$ have the same truth value in the collapsed model.

However, it is easy enough to construct a theory in an extended language, which contains N_n and which does have a truth predicate in this sense. (The details are as in Dowden (1984).)

To formulate the sentences involved in the other results we need a proof predicate for N_n. The simplest thing to do is to take the predicate B that defines N_n in N_n. (vii) and (viii) are given to us straight away. (For (viii) the decision procedure will give us a proof of $\neg B(<\varphi>)$, and decision procedures are certainly finitary.) For (ix), standard techniques allow us to construct a Gödel sentence, ψ, itself of the form $\neg B(<\psi>)$. If ψ is provable, all well and good. If it is not, $\neg B(<\psi>)$ is provable so ψ is provable anyway.

Alternatively, we could take a B satisfying the identity condition. For such a B, if φ is not provable, $B<\varphi>$ is not provable, and so its negation is. This suffices for the other results.

Appendix 2

In this appendix, I will give a formalisation of Petersen's argument, used in §4. The argument is carried out in the language of arithmetic, augmented by a least number operator, μ, and a two-place denotation predicate, Δ. Given a formula of one free variable, x, φ, "$\mu x\varphi$" refers to the least number satisfying φ, if there is one, or 0 otherwise. It therefore satisfies the following description principle:

$$\exists x\varphi \rightarrow \varphi(x/\mu x\varphi)$$

Δ satisfies the following two conditions:

$$\Delta(<t>,t) \qquad (\Delta 1)$$

$$\Delta(x,y) \wedge \Delta(x,z) \rightarrow y = z \qquad (\Delta 2)$$

where t is any closed term. Between them, $\Delta 1$ and $\Delta 2$ say that $<t>$ refers to t and only t. (This is reasonable since, as μ is defined, all terms denote.)

For the proof, we need to have an appropriate way of representing self-reference. This is done with a version of Gödel's fixed point lemma. We will assume that every one-place primitive recursive function is represented by a term of the language—in fact, just diagonalisation will suffice. (This is not essential, but simplifies the proof.) The lemma is as follows:

Lemma

For any term with one free variable, $\tau(x)$, there is a term, σ, such that we can prove:

$$\sigma=\tau(<\sigma>)$$

Proof

If t is any term, let its diagonalisation be the term obtained by substituting $<t>$ for each free variable in t. By assumption, there is a term, $\delta(x)$, that represents diagonalisation, i.e., if s is the diagonalisation of t then we can prove that $\delta(<t>)=<s>$. Now consider $\tau(\delta(x))$. Suppose this has code m. Then its diagonalisation is $\tau(\delta(\mathbf{m}))$. Suppose that this had code n. Then since δ represents diagonalisation, we can prove:

$$\delta(\mathbf{m})=\mathbf{n}$$

And hence

$$\tau(\delta(\mathbf{m}))=\tau(\mathbf{n})$$

Taking σ to be $\tau(\delta(\mathbf{m}))$ gives the result. □

The formalisation of Petersen's argument is now simple. Consider the term $\mu x\Delta(y,x)+1$. By the diagonal lemma we can find a term, π such that:

$$\pi=\mu x\Delta(<\pi>,x)+1 \qquad (1)$$

$\Delta 1$ gives us:

$$\Delta(<\pi>, \pi) \qquad (2)$$

Hence, $\exists x\Delta(<\pi>,x)$. By the description principle:

$$\Delta(<\pi>, \mu x\Delta(<\pi>,x)) \qquad (3)$$

But then by (2), (3) and $\Delta 2$:

$$\pi=\mu x\Delta(<\pi>,x)$$

So by (1), $\pi=\pi+1$; and hence, $\exists x x=x+1$.

Department of Philosophy
University of Queensland
Queensland 4072
Australia

GRAHAM PRIEST

REFERENCES

van Bendegem, J. P. 1991: "Strict, Yet Rich Finitism". Typescript.

Boolos, G. and Jeffrey, R. 1974: *Computability and Logic*. Cambridge: Cambridge University Press.

Diamond, C. 1976: *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge 1939*. London: Harvester Press.

Dowden, B. 1984: "Accepting Inconsistencies from the Paradoxes". *Journal of Philosophical Logic*, 13, pp. 125-30.

Fraenkel, A., Bar-Hillel, Y., and Levy, A. 1973: *Foundations of Set Theory*. (2nd ed.). Amsterdam: North-Holland.

Hilbert, D. 1925: "On the Infinite". *Mathematische Annalen*, 95, pp. 161-90. Reprinted in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings*. Oxford: Blackwell, 1964.

Kripke, S. 1982: *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Lauden, L. 1977: *Progress and its Problems: Towards a Theory of Scientific Growth*. London: Routledge & Kegan Paul.

Lewis, D. 1984: "Putnam's Paradox". *Australasian Journal of Philosophy*, 62, pp. 221-36.

Meyer, R. 1976: "Relevant Arithmetic". *Bulletin of the Section of Logic, Polish Academy of Sciences*, 5, pp. 133-7.

Meyer, R. and Mortensen 1984: "Inconsistent Models for Relevant Arithmetics". *Journal of Symbolic Logic*, 49, pp. 917-29.

Priest, G. 1983: "Logical Paradoxes and the Law of Excluded Middle". *Philosophical Quarterly*, 33, pp. 160-5.

——1987: *In Contradiction*. The Hague: Martinus Nijhoff.

——1991: "Minimally Inconsistent LP". *Studia Logica*, 50, pp. 321-31.

——1994: "Creativity and Gödel's Theorem", Ch. 6 of T. Dartnall (ed.), *Creativity*. Dordrecht: Kluwer Academic Publishers.

Putnam, H. 1983: "Models and Reality", in *Realism and Reason,* Philosophical Papers Vol. 3, pp. 1-25. Cambridge: Cambridge University Press.

Wright, C. 1985: "Skolem and the Skeptic, II". P*roceedings of the Aristotelian Society*, Supplementary Volume 59, pp. 118-37.