ABOUT "THE PHILOSOPHICAL SIGNIFICANCE OF GÖDEL'S THEOREM": SOME ISSUES

Crispin Wright

1. DUMMETT'S PROBLEM

It is a very natural supposition that, for any particular consistent formal system of arithmetic, one of the pair consisting of the Gödel sentence and its negation must be true. This was rejected by Wittgenstein in the notorious appendix on Gödel's Theorem in the *Remarks on the Foundations of Mathematics.*[1] Wittgenstein there implicitly repudiated not merely any Platonist conception of mathematics, as usually conceived, but the much more deeply rooted idea that arithmetic is in the business of *description* of a proper subject matter of any kind. His view, it seems, was that there simply is no defensible conception of truth for the sentences of a formal arithmetic which might coherently — whether or not justifiably — be thought to outrun derivability within it.

Wittgenstein's stance here goes right against the grain. Whatever the situation in other areas of pure mathematics, formal number theory, we intuitively feel, has to answer to a very definite pre-formal conception of the natural numbers, their structure and basic properties. We don't come by our basic arithmetical concepts by doing formal number theory; and the informal understanding which we acquire of those concepts suffices not merely to give the Peano axioms a cogency which would be quite missing if they had anything of the character of stipulations, but also to supply, so it seems, a definite interpretation, at least in principle, for any arithmetical sentence. Thus it is hard not to accept that there is a subject matter to which the Peano axioms answer, and which, since no material vagueness intrudes, must confer determinate truth-values on a Gödel sentence and its negation.

This view of the matter, however, sets up a prima facie difficulty — the

[1] See [171], Part I, Appendix III, pp. 116–123.

B. McGuinness and G. Oliveri (eds.), The Philosophy of Michael Dummett, 167–202.

difficulty to which Michael Dummett's well-known discussion responds.[2] As soon as it is granted that any intuitively sound system of arithmetic merely *partially* describes the subject matter to which it answers, an explanation is owing of *how* the subject matter in question can possess a determinacy transcending complete description. No doubt there are species of Platonism for which this demand would pose no special embarrassment. On such *rampant* Platonist views (John McDowell's term), description-transcendent determinacy should be no more puzzling, in the case of the subject matter of arithmetic, than it is in the case of the subject matter of physics, say, or astronomy — or in any case where a domain of reality is constituted in full independence of the conceptualisations and investigations performed by human beings. But any attractive form of Platonism will have to take a more moderate, 'conceptualist' line: the subject matter of arithmetic, and mathematics generally, whatever autonomy it may go on to acquire, will be conceived as originally constituted in the *arithmetical understanding* — so that the special concern of the arithmetician will be with the implications of our arithmetical concepts as properly understood.

It is this saner Platonism which seems open to the advertised difficulty. The problem, simply, is that to allow that the subject matter of arithmetic is determinate beyond any systematic description is apparently, for the moderate Platonist, to admit that our understanding of arithmetical concepts transcends any systematic description of correct arithmetical practice: that no matter what we offer as a characterisation of what is non-contingently true of the natural numbers, we will leave the status of certain such truths indeterminate. And that, it seems, is exactly for meaning to transcend use, just as Dummett says. The subject matter of arithmetic, on this view, is constituted by a definite conception which we hold 'in mind'; but any attempt at a systematic characterisation of that conception will fail to determine the truth-values of statements on which the conception itself is not neutral.

Outlined in this way, the problem depends on an intuitive realism about arithmetic, with a bivalent conception of arithmetical truth as the principal ingredient. And there are alternatives to this realism besides the radical anti-descriptivism of Wittgenstein's Appendix. Surely it is possible to think of arithmetic as having a proper subject matter,

[2]M. Dummett, 'The philosophical significance of Gödel's Theorem', in [50], pp. 186–201. Originally published in *Ratio* Vol. no. 5, 1963, pp. 140–155.

which differing formal systems might portray with varying degrees of completeness, without being railroaded into an acceptance of the principle of Bivalence for arithmetical statements generally? Indeed, just that is the characteristic attitude of Mathematical Intuitionism. Might not the problem be avoided altogether on such a view? Once we discard the notion that every well-formed arithmetical sentence must be determinately true or false, there is no longer any reason to suppose that any particular arithmetical system must, by omitting a Gödel sentence and its negation, omit a determinate arithmetical truth.

As Dummett sets up the problem, however, discarding Bivalence would be of no avail. This is because:[3]

> By Gödel's theorem there exists, for any intuitively correct formal system for elementary arithmetic, a statement U expressible in the system but not provable in it, which not only is true but can be recognised by us to be true.

That is Dummett's opening statement of Gödel's 'result'. Like most such statements in the literature, it adds something on which Gödel's mathematical work is strictly neutral: not an endorsement of a bivalent conception of arithmetical truth, but the contention that *we are able to recognise the truth* of the universally quantified Gödel sentence. The idea is, indeed, quite orthodox that to follow Gödel's reasoning is, in the light of an ordinary understanding of arithmetical notions, to see that U — the universally quantified undecidable sentence — although unprovable in the system in question if the latter is consistent, is nevertheless an arithmetical truth. This thought involves no commitment to the principle of Bivalence, and could at the same time, I suppose, be repudiated by one who accepted Bivalence for arithmetic. But it apparently sets up just the same problem. If it is right, then, as before, it seems that no matter what systematic characterisation we give of the range of statements which we understand to be true of the natural numbers, that characterisation will be neutral on cases on which the understanding itself is not — cases which, following Gödel, we may informally recognise to be true.

Dummett offers a solution to the problem which exploits an extremely important idea he has called *indefinite extensibility*. A concept is indefinitely extensible if, roughly, any attempt at a comprehensive characterisation of its extension immediately subserves the specification of new

[3] See [39], p. 186.

instances of the concept which that characterisation demonstrably omits. A case can be made for regarding each of *class, ordinal number* and *real number* as indefinitely extensible in this way. And the real significance of Gödel's result, Dummett contends, is that it shows that the concepts, *statement true of all the natural numbers,* and *ground for affirming that all natural numbers have a certain property* are in like case. A solution to the problem is then to be accomplished, if I have not misunderstood, by the reflection that indefinite extensibility is a kind of *vagueness*, so that accepting the statement U as a further arithmetical truth need not be a fully determinate obligation of the arithmetical understanding — an understanding which, since the Gödelian construction may be iterated indefinitely, we would otherwise have to regard as transcending any systematic description of the obligations it imposed.

I have some misgivings, not about the idea of indefinite extensibility in general, but about its application in this context and its construal as a kind of vagueness. I will briefly return to Dummett's solution to his problem at the end of the paper.

2. LUCAS AND PENROSE AGAINST MECHANISM

The problem concerning meaning and use is one of two major general philosophical concerns on which Gödel's theorem has been thought to bear. But the other has attracted a much greater volume of commentary. Writers such as J.R. Lucas[4] and, more recently, Roger Penrose[5] have argued that Gödel's theorem provides as clear a demonstration as philosophy could reasonably hope for that our arithmetical capacities in particular, and hence the powers of the human intellect in general, cannot in principle be simulated by a machine. Here is an expression of the idea by Penrose:[6]

> ... *whatever* ... algorithm a mathematician might use to establish mathematical truth ... there will always be mathematical propositions, such as the explicit Gödel proposition [for the formal system associated with that algorithm], that his algorithm cannot provide an answer for. If the workings of the mathe-

---

[4] See [99], pp. 112–137.

[5] See [131]. See especially the section entitled 'The non-algorithmic nature of mathematical insight' in chapter 10, pp. 538–541. See also the sections, 'Gödel's theorem' and 'Mathematical insight' in chapter 4, pp. 138–146.

[6] See [131], pp. 538–539.

> matician's mind are entirely algorithmic, then the algorithm ...that he actually uses to form his judgements is not capable of dealing with the [Gödelian] proposition ...constructed from his personal algorithm. Nevertheless *we* can (in principle) see that [that proposition] is *true*. That would seem to provide *him* with a contradiction, since *he* ought to be able to see that also. Perhaps this indicates that the mathematician was *not* using an algorithm at all!

The central thought here is very simple: Gödel's theorem teaches us that the class of arithmetical truths ratifiable by the human mind does not coincide with those deliverable by any particular program for proof construction — any particular algorithm. Yet anything worth regarding as a *machine* has to be slave to such an algorithm. It follows that our arithmetical powers are not exhaustively mechanical.

A very distinguished tradition of disbelief has grown up in response to this line of argument. One common objection[7] is that our ability to construct a Gödel sentence for a particular formal system, and hence to recognise the truth of that sentence, is of course hostage to our capacity intelligently to receive a specification of the formal system — algorithm — in question. The Lucas/Penrose argument, the objection goes, simply overlooks the possibility that human arithmetical capacity is indeed encoded in a particular formal system of which, however, we are unable to comprehend any formal specification suitable for an application of the Gödelian construction. It need not be questioned that we can produce a Gödel sentence for any arithmetical system which we can take in. But even granting that we can recognise such a sentence to be true, what follows is merely a disjunction: *either* no such system encodes all human arithmetical capacity — the Lucas/Penrose thought — *or* any system which does has no axiomatic specification which human beings can comprehend.

Gödel himself seems to have proposed essentially this conclusion. He writes:[8]

> For if the human mind were equivalent to a finite machine ...there would exist ***absolutely*** unsolvable Diophantine prob-

[7] See for instance [9].

[8] From ms. page 13 of Gödel's 1951 Josiah Willard Gibbs Lecture, delivered to the American Mathematical Society. See [71], volume III. (The volume will be preserving the page numbers of the typescripts of Gödel's in the margins.) For discussion of Gödel's view, see G. Boolos's *Introductory Note to # 1951*, Ibid.

> lems of the type described above, where the epithet 'absolutely' means that they would be undecidable, not just within some particular axiomatic system, but by *any* mathematical proof the human mind can conceive. So the following disjunctive conclusion is inevitable: ...*the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable Diophantine problems of the type specified.*

But is this conservative — disjunctive — conclusion the most that is justified? It is never very clear what is meant by talk of the powers of the 'human mind'. But in this context the unclarity is urgent. If we are talking about actual human minds, constrained as they are by the volume and powers of actual human brains, then it can hardly be doubted that the range of arithmetical sentences whose truth we are capable, by whatever means, of ratifying, is finite. So there are, of course, algorithms which capture all those sentences and more. Since Lucas and Penrose are presumably in no doubt about that, we should conclude that the anti-mechanist thesis concerns the character of mathematical thought as reflected in what *in principle* we have the power to accomplish in mathematics. The comparison has to be not between *actual* human arithmetical capacities, even those of the most gifted and prodigious mathematicians, and *actual* theorem-proving machines, even the most powerful and sophisticated that technology will ever accomplish, but between what we can do in principle and what in principle can be accomplished by a machine, where a 'machine' is constrained only to the extent that its theorem-proving capacities correspond to the output of an effectively axiomatised system.

But what ought 'in principle' to mean for the purposes of such a comparison? I propose we should count any feat as within our compass in principle just in case some *finite* extension of powers we actually have would enable us to accomplish it *in practice*. Correspondingly for the powers in principle of a theorem-proving machine. So each first-order deductive consequence of the axioms which it encodes can in principle be proved by such a machine. Likewise, any arithmetical statement is ratifiable in principle by the human mind just in case, under some finite extension of capacities we actually possess, we could ratify it in practice.

Well plainly, if these are the terms of the intended comparison, then the mooted reservation about the Lucas/Penrose line lapses. There is no need for the second disjunct in the Gödelian statement. For any effectively axiomatised system of arithmetic is, in the relevant sense, *in principle surveyable* by human beings, who may consequently, in principle, carry out the Gödelian construction upon it. Even if in practice, limited as we actually are, our actual arithmetical capacities can be captured by a formal system whose specification we can neither comprehend nor, therefore, Gödelise, it remains that we have in principle the resources — a finite extension of our capacities would enable us — to do so. The qualification marked by the second disjunct in Gödel's statement is called for only if the words 'can conceive' concern human abilities in practice.

This is likely, I imagine, to provoke the following complaint. Let it be, for the sake of argument, that, in the sense of 'in principle' stipulated, what the human mind can do in principle and what is in principle possible for a theorem-proving machine do indeed diverge. Still, the intention of the opponents of mechanism is surely to make good a claim which concerns not *semi-divine* creatures — creatures whose powers in practice coincide with ours in principle — but actual living human beings. We were supposed to learn something about how to regard *ourselves*, limited as we are, and not about supermen who differ from us by having no finite limits of memory, rate of working, concentration, intellectual fluency, and so on.

However, I believe it is just a mistake to think that if the claimed disanalogy in powers 'in principle' were sustained, we would still learn nothing about our actual limited selves. On the contrary, we would learn that, limited though we are, we work with a concept of arithmetical demonstration whose extension does not allow of effective enumeration. By contrast, the extension of the concept of demonstration encoded in any formal system that might serve as the program for a theorem-proving machine *is* effectively enumerable. A satisfactory description of our actual arithmetical powers would thus have to acknowledge that they are informed by a concept of a sort by which, if our arithmetical thought were purely the implementation, as it were, of an in principle specifiable formal system, they would not be informed. Hence, if a disanalogy has been pointed to at all, then it indeed is between human beings as we actually are and as we would be if our mathematical thought were entirely algorithmic. The concept of demonstration which actually informs our construction and ratification of proofs would have been shown to have a feature which it would lack if our powers in those respects were purely mechanical.

If this is right, then there is a connection, seldom remarked on, between the two most general philosophical debates on which Gödel's theorem has been thought to bear. Dummett's problem, the prima facie difficulty raised by the theorem for the idea that meaning is determined by use, only arises if we take it that the theorem impels us to recognise that the extension of the concept, 'true of all natural numbers', outruns any attempt formally to characterise the conditions under which it is properly applied. And unless we accept that arithmetical statements are unrestrictedly bivalent, we are so impelled only if, with Dummett, we think that an informal demonstration of U attends the Gödelian construction. But if that is true — if Gödel's theorem teaches us that arithmetical truth outruns any systematic characterisation of its extension precisely by providing the basis for an informal demonstration of the universally closed undecidable sentence — then we simultaneously learn that our concept of an intuitively correct arithmetical demonstration cannot be thought of as the concept of whatever is sanctioned by some in principle specifiable set of algorithmic procedures. And that, Lucas and Penrose should contend, is just what mechanism, if it is to aspire to any clarity, must deny. So unless we accept Bivalence, Dummett's problem arises only on a supposition whose correctness would promise to settle the other debate in favour of Lucas and Penrose.

But of course the most prevalent line of criticism of the Lucas/Penrose thesis has been precisely directed at the supposition that we can 'see' the universally quantified undecidable sentence (U) to be true — that Gödel's construction somehow provides or underpins an informal demonstration of U in the first place. Lucas and Penrose, the charge is, conveniently forget that any ground which emerges from Gödel's reasoning for accepting the truth of U is *entirely dependent* on the hypothesis of the consistency of the object system. And for that hypothesis Gödel's reasoning produces, of course, no grounds whatever. It may be a hypothesis of whose truth we are confident; and perhaps we are entitled to be so. Whatever sort of entitlement that may be then transfers into an entitlement to accept the truth of U. But that's a far cry from saying that something worth regarding as a mathematical *demonstration* of U is in the offing. Hilary Putnam writes:[9]

[9] See [135], in [134], p. 366.

> Let T be the Turing machine which 'represents' me in the sense that T can prove just the mathematical statements I can prove. Then the argument ... is that by using Gödel's technique I can discover a proposition that T cannot prove, and moreover I can prove this proposition. This refutes the assumption that T 'represents' me, hence I am not a Turing machine. The fallacy is a misapplication of Gödel's theorem, pure and simple. Given an arbitrary machine T, all I can do is find a proposition U such that I can prove:
>
> (3) If T is consistent, U is true,
>
> where U is undecidable by T if T is in fact consistent. However, T can perfectly well prove (3) too! And the statement U, which T *cannot* prove (assuming consistency), I cannot prove either (unless I can prove that T is consistent, which is unlikely if T is very complicated.)

On this view, Gödel's construction ought to be regarded as providing no reason to affirm that the concepts, 'demonstration that all natural numbers have a certain property', or 'statement true of all the natural numbers' are indefinitely extensible. Critics of Lucas and Penrose who advance this line ought therefore to hold that Dummett's problem is likewise misconceived. But are they right?

3. DUMMETT'S MODUS PONENS

Why did Dummett himself accept that U 'can be recognised by us to be true'? He writes:[10]

> The argument for the truth of U proceeds under the hypothesis that the formal system in question is consistent.

and then again:[11]

[10] See [39], p. 192.
[11] See [39], p. 194.

> The argument to establish the truth of U involves establishing the consistency of the formal system. The interest of Gödel's theorem lies in its applicability to *any* intuitively correct system for number theory... We have, therefore, to consider the consistency proof with which Gödel's reasoning must be supplemented if the truth of U is to be established as one we know we can give for *any* formal system [my emphasis], provided only that it is assumed about that system that it is intuitively correct.

These remarks make it clear that the general form of demonstration of the Gödel sentence envisaged by Dummett is in effect a modus ponens on Putnam's conditional (3) — a conditional which, as Putnam remarked, can be established *within* whatever formal system Gödel's construction is applied to. Dummett's idea seems merely to be that, provided only that an arithmetic-containing system is intuitively correct, we can always accomplish a proof that it is consistent and then detach the consequent of the appropriate instance of (3).

Taken in full generality, that claim may seem merely preposterous — surely we may have no idea how to establish the consistency of a complex but intuitively sound system of set-theory, for instance. But as Dummett enlarges on his claim, it is not controversial at all. Provided we are given that the axioms of the system are true, and that its rules of inference are truth preserving, a routine induction on the length of proofs in the system will, as Dummett remarks, yield that all its theorems are true and hence, since no true statement has a true (syntactic) negation, that of no statement can it be the case that both it and its negation are theorems. Dummett acknowledges, of course, that a proof of this kind is hardly 'genuinely informative' (his phrase). A 'genuinely informative' consistency proof must (presumably) do more than elicit, in this trivial way, an implication of the presupposition that the axioms of the system are true and its rules of inference sound. It is additionally required, I take it, to afford some positive reason for regarding the intuitive credibility of the axioms and rules as an indication that they *are* indeed respectively true and sound. Thus the sort of consistency proof we would ideally want for an intuitively satisfactory system of arithmetic would be one which, rather than merely drawing out the implications of the presupposition that the intuitive cogency of the axioms is an indication of their truth, provided persuasive reason for reposing confidence in arithmetical intuition in the first place.

The kind of consistency proof to which Dummett adverts is powerless to do that.[12] But it is only a proof of that uninformative sort that there is reason to think can be obtained in every case where we have an intuitively correct system of arithmetic.[13] So how should we appraise the claim that Dummett's modus ponens *demonstrates* U? Obviously, it depends on what, in this context, is required of a demonstration. Here is one natural notion: an intellectual routine counts as a demonstration of P just in case an agnostic about P could nevertheless perfectly reasonably place confidence in the methods and principles deployed in the

[12]Such a 'demonstration' of consistency would have been available for Naive Set Theory before it was known that the naive Abstraction axiom was inconsistent.

[13]It is a nice question whether Gödel's second incompleteness theorem, that no consistent system of arithmetic can contain the proof-theoretic resources for a demonstration of its own consistency, provides the basis of a case that nothing better than an uninformative consistency proof can be obtained for *any* system of arithmetic to which the first theorem applies. The question is, in particular, whether the non-arithmetical resources deployed in formal proofs of the consistency of arithmetical systems, for example the use made of transfinite induction up to ε_0 in the famous Gentzen proof, are essentially no more transparent — no more epistemically secure — than the proof theory, and especially the full-fledged Induction schema, of first-order number theory itself. Gentzen himself thought otherwise:

> ...I am inclined to believe that in terms of the fundamental distinction between disputable and indisputable methods of proof, the *proof of the finiteness* of the reduction procedure

the part of his proof that uses transfinite induction

> can still be considered indisputable, so that the consistency proof represents a *real vindication* of the disputable parts of elementary number theory.

(From [67], chapter 4, p. 197.) The issue is certainly subtle and deserves detailed discussion. But it is hard not to feel sceptical about Gentzen's view. To accept the Gentzen proof is to be persuaded of a mapping between the proofs constructible in elementary number theory and the series of ordinals up to ε_0. And to understand the structure of the ordinals up to ε_0 is to grasp a concept which embeds and *builds on* an ordinary understanding of the series of natural numbers. So to trust the Gentzen proof is implicitly to forgo any doubt about the coherence of the concept of natural number. Any residual doubt about the consistency of first-order number theory would have, therefore, to concern whether the (fully coherent) concept of number is faithfully reflected by the standard first-order axioms — specifically, by all admissible instances of the induction schema. Such a doubt would have to concern whether, even when restricted to first-order arithmetical vocabulary, 'tricksy' predicates might not somehow be formulated whose use in inductions could lead to contradiction. Could someone reasonably worry about that who was confident in the consistency of Gentzen's methods? How might such a doubt be further elaborated?

routine, and could arrive, on the basis of following it through, at considerations which would rationally oblige him a priori to assent to P. Say that a proof is *suasive* if it meets those conditions. Now clearly, since it takes the truth of the axioms and the soundness of the underlying logic as a premise, the sort of consistency proof Dummett had in mind can hardly be suasive — (except perhaps for a rather dim thinker to whom it has not occurred that you cannot get contradictions in a system with true axioms and truth-preserving rules.) So it furnishes no demonstration of U in the stipulated sense; in particular, it has nothing to offer someone who already accepts, as it were on faith, the consistency of the system in question but now wants a mathematical argument that he is right to do so — something which he could use, for instance, to respond to a critic.

Suppose, on the other hand, we drop the requirement of suasiveness. Then the sort of non-suasive demonstration of the consistency of S outlined doesn't contribute to something worth describing as, in Dummett's words, an 'argument to establish the truth of U'. In particular, it has nothing to offer a thinker who is agnostic about U and accepts the consistency of S only as a working assumption. Accordingly no ground is provided by such a demonstration for the presupposition of Dummett's problem, that the class of statements which we can recognise to be true of the natural numbers outruns any systematic characterisation.

The dilemma, then, is this. If a demonstration has to be suasive, the construction of the Dummettian modus ponens will not in general constitute a demonstration and nothing obstructs Putnam's claim that there is, as a general rule, no demonstrating the Gödel sentence. But if demonstration doesn't have to be suasive, then while the Dummettian modus ponens can in general rank as a demonstration, the demonstrability of U will provide no basis for Dummett's problem or, indeed, for the contention of Lucas and Penrose.

4. EMPIRICALLY BASED CONSISTENCY

There is, however, a disanalogy. The Lucas/Penrose line requires, of course, that there be a demonstration — a cogent *a priori* argument — for the truth of the Gödel sentence. Their case entirely depends on the putative contrast between the concept of proof operated by the human mathematician and the concept of proof that governs the operations of a particular Turing machine. By contrast, would it not suffice to set up

Dummett's problem if there were, not a priori, but *empirical-inductive* grounds for the assertion of the Gödel sentence? In order for the concept, 'Statement which may justifiably be claimed to be true of all the natural numbers', to be disclosed as indefinitely extensible by Gödel's result, it ought to suffice that we can invariably in principle produce a Gödel sentence for a given formal arithmetic, and then produce grounds *of some sort* for accepting it as true. If we have empirical-inductive reason to believe in the consistency of a particular arithmetic, this will transfer to the Gödel sentence which, by the conditional (3), may then be inferred. And if we had reason to believe not merely in the consistency but in the *truth* of the original system, then we will thereby have reason to believe in the truth, and hence the consistency, of that system augmented by its Gödel sentence. So the process will iterate.

But *do* we have empirical-inductive reason to believe in the consistency of arithmetic, or in that of any other branch of mathematics where an 'informative' consistency proof is not to be had? Hartry Field writes:[14]

> ...a large part of the *reason* most of us believe that modern set theory is consistent is the thought that if it weren't consistent someone would probably have discovered an inconsistency by now.

Likewise:[15]

> ...much of our knowledge of possibility is to some extent inductive. For instance, our knowledge that [von Neumann-Bernays-Gödel set theory is consistent] seems to be based in part on the fact that we have been unable to find any inconsistency in [that system].

Obviously enough, however, there cannot be a *strictly* inductive grounding for beliefs in consistency in such cases. Such grounding would beg an established correlation between the availability of the relevant kind of ground — viz. no-one's yet having found any inconsistency — and the relevant system's actually *being* consistent. There would therefore have to be independent grounds for the latter — which is just what we do not have. Field's thought has to be, rather that there is here an appropriate inference to the best explanation (or something of that sort). Roughly: things have reached a pass where the *most likely* explanation

[14] See [62], p. 232.
[15] See [62], p. 88.

of the failure of any inconsistency to turn up is that there is none to turn up.

Why though, is that the favoured explanation — what gives us our grip on the probabilities? Why is it supposed to be relatively unlikely that, in at least some such cases, inconsistencies do exist to be unearthed, but by tricks and turns more subtle than any that have occurred to us? In particular, if we say that it is *improbable* that that is the situation of ordinary arithmetic, then again: on what basis are the probabilities established? If someone claims that, for reasons such as these, Field's idea is simply a mistake, I do not think it is at all clear how to refute them. And unless the claim can be refuted, it may begin to seem not merely that the *modus ponens* route to the truth of the Gödel sentence founders for want of any real ground, a priori or empirical-inductive, for its minor premise, but more: that our faith in the consistency of the arithmetical part of our mathematical thinking, and hence in all that builds upon it, is without any justification whatever.[16]

[16]The status of this faith is an extremely interesting epistemological issue, well worth detailed investigation. Unfortunately, that must lie beyond the scope of this discussion. But one immediate consideration about the sceptical turn just bruited is that it wholly ignores the epistemological territory which was the centre of attention in Wittgenstein's notes *On Certainty*(See [164]). Very crudely: beliefs for which one has no *earned* justification need not be beliefs for which one has no justification *tout court*. All justification, Wittgenstein argues, takes place within some system of beliefs: a system in which an unexamined heritage of belief will be a major component, but in which there will also be certain very general empirical-seeming beliefs whose logical role places them *beneath* evidential support (cf. [177]), as well as fundamental principles of inference — principles of inference which we find intuitively completely compelling but of which no suasive demonstration can be constructed. A leading theme in *On Certainty* is that in cases of each of these kinds the justifiability of a belief need not depend on any cognitive *accomplishment* specific to the proposition concerned.

It is a nice question whether or in what directions a compelling development of this idea may be made explicit. But it may be thought to strike an intuitive chord as far as the consistency of elementary arithmetic is concerned. Many would feel, that is to say, that the soundness of our arithmetical thought is something which we are entitled to assume on no other basis than its intuitive cogency, and that a demonstration which presupposes it as a premise is no more precarious on that account than any intuitively cogent demonstration. At the very least, it should be granted that the unavailability of both suasive a priori demonstration and empirical confirmation of the consistency of suitable systems of arithmetic does at least not *immediately* write off the Dummettian modus ponens as a suasive demonstration of U. The strategy remains of trying to make out that the belief in the consistency of intuitively sound systems of arithmetic is one in which we are *groundlessly* justified. But I will not pursue the issue further here.

However, it is not obvious that any demonstration of the Gödel sentence lurking in the vicinity *has* to be a Dummettian modus ponens. I shall outline two further trains of thought which attempt to proceed differently.

5. THE SIMPLE ARGUMENT

The first is, at first blush, quite straightforward. Let S be an intuitively acceptable system of arithmetic — say, standard first order Peano arithmetic — and U the universally closed Gödel sentence for S. Then we have a demonstration of

(3) if S is consistent, then U,

and hence of

(3′) if S is sound, then U,

where the soundness of S consists in the truth of its axioms and the validity of its principles of inference. Now consider any formal derivation, D, of a sentence, P, carried out in S. S is an intuitively acceptable system of arithmetic, and so we would ordinarily regard D as entitling us to affirm P unconditionally. But there are, of course, implicit conditions. Someone who regards the derivability of a particular sentence within a particular formal system as justifying the assertion of that sentence under some customary interpretation presupposes that the system in question is *faithful* to that interpretation: that the inference rules involved are valid, and that the axioms express truths when so interpreted. Thus, the contention of the Simple Argument is, what is immediately justified by D is only the *conditional* claim that provided S is sound (under the intended interpretation), then P is true.

The claim, then, is that *any* proof carried out within S strictly entitles one to no more than the conclusion:

If S is sound, then P,

and that in regarding such a proof as a demonstration of P *tout court*, we merely — justifiably enough — suppress the standard assumption of soundness. If that is right, then the thought may be encouraged that the conditional (3′), which it is agreed on all hands is established by Gödel's reasoning, differs in no interesting way from the form of what, strictly, is established by *any* derivation within an intuitively acceptable formal arithmetic. Mere consistency of policy thus demands, so says the argument, that we recognise Gödel's reasoning as a proof of U. True, all that is explicitly proved is the conditional (3), and hence (3′). But the warrant to affirm a sentence as demonstrated in arithmetic is *always* conditional on an assumption of soundness. That warrant can hardly be weakened if such an assumption, rather than playing its customary implicit role, actually figures explicitly, as in (3′).

This train of thought needs one immediate qualification. The claim of any proof within a formal arithmetic to rank as a demonstration of its conclusion need not depend upon the supposition of the soundness of the system as a whole. It will suffice if the *particular* axioms and rules of inference deployed in the proof are sound, and these need not of course comprise all the axioms and rules which the system involves. In this way, then, the standing of a proof within S as a demonstration may turn on less than the supposition that S is sound overall. So the strongest conclusion in prospect, if this line of thought is sustained, will be that the status of U in point of credibility, after we have the demonstration of (3) and hence (3′), is comparable to that of any sentence P, of which we have a syntactically correct proof within S, which appeals to *all* the axioms and deductive resources of S.[17] Since we might have more confidence in some aspects of S than others, the credibility given to U by the demonstration of (3) may not be as great as that afforded by some formal proofs within S. Still, this seems no very damaging a limitation.

The idea of the Simple Argument, then, is that, by contrast with the Dummettian modus ponens, a demonstration of U is effectively provided just by the formal proof of (3). There is no need to make a special case for the antecedent, viz. the consistency of S, since when S is an intuitively acceptable system of arithmetic — the only case which interests us — the assumption of the soundness of S is anyway implicit in the whole practice of receiving its formal proofs as demonstrations. Either, then, we should accept the proof of (3), hence (3′), as a demonstration of U, or we should cease to regard formal proofs constructible within S — or at least those which make maximal use of its resources — as demonstrations in the first place.

[17] Cf. [39], pp. 192–193.

This line can seem quite arresting. But can it really work? Grant that the soundness of the principles employed is presupposed when we treat a formal proof as a demonstration, and remains as an implicit proviso when we proceed to affirm its conclusion. The question is whether it follows that something of the form

(4) If (the relevant part of) S is sound, then P

is the strongest claim we are ever 'strictly' entitled to. Why, for instance, stop there? Will not the verification of such a conditional itself be the result of some sort of substantial cognitive process? And does not our confidence in the conditional's truth implicitly assume, therefore, the soundness of any principles appealed to in that process and the proper functioning of all relevant faculties? So are we not 'strictly' entitled merely to something of the form,

> If [those principles are sound, and all relevant faculties functioned properly, etc.], then if (the relevant part of) S is sound, then P?

If so, where does this attenuation of strict entitlement stop? How are we ever 'strictly' entitled to anything? If not, what justifies the initial claim that (4) exhausts our strict entitlement? The moral suggests itself: granted that we implicitly presuppose soundness whenever we treat a formal proof as establishing a conclusion, P, it is nevertheless simply *incoherent to equate the role played by that presupposition with that of the tacitly discharged antecedent of a conditional whose consequent is P.*

That is one objection. But there is a strong suspicion that another, independent confusion is at work in the Simple Argument. If S is sound, then a formal proof in S of P really does demonstrate that P is true, and by running through the proof we really do *recognise* its truth. (Compare: if I am not now dreaming, then my visual experiences really do constitute a perception of my hand in front of my face and, by having those experiences, I really do recognise that my hand is there.) Presented with such a proof, therefore, we should be at least entitled to

> (4′) If (the relevant part of) S is sound, then P has been demonstrated.

which of course entails but is not entailed by (4). Given (4′), the soundness of the relevant part of S is all we need to ensure that our running through a certain intellectual routine constitutes our actually verifying the truth of P. But what routine constitutes such a verification if all we are given is (4)? The only candidate is the proof of (4) itself, along with the thought that the truth of its antecedent is a presupposition of a kind we make in treating any proof in S as a demonstration. But *that* thought only makes the case so long as it is supposed to follow that conditions like (4) are all we ever strictly get. And even if it is granted that all we ever 'strictly' get are conditionals with such antecedents, it remains that they are, in the arguably crucial respect illustrated by (4′), always stronger than those epitomised by (4). That a routine's claim to demonstrate P always rests on certain presuppositions is one thing; it does not follow that we cannot ask more of a demonstration of P than to be supplied with a proof that, if those presuppositions are true, then so is P.

It does not follow, but perhaps it might independently be argued. I do not claim that either of these lines of objection is decisive, only that the Simple Argument is not so simple as it seems. But a proponent of the view that the proof of (3) is, in effect, a demonstration of U, clearly has some work to do.

6. AN INTUITIONISTIC DEMONSTRATION?

Dummett's modus ponens and something like the Simple Argument have each, I suspect, been at work, though in various degrees of explicitness, in the thought of many who have been inclined to regard U as effectively demonstrated. But there is also a, so far as I am aware, unremarked case for saying that a specific *intuitionistic* demonstration of U arises directly out of Gödel's own reasoning. This demonstration does not proceed, like Dummett's modus ponens, from an undischarged hypothesis of the consistency of S (or from something stronger); but nor, like the Simple Argument, is it content to allow that Putnam's conditional (3) is the most that can strictly be demonstrated. The key ingredients are the intuitionists' official account of negation and the reflection that Gödel provides the resources for finding a contradiction in an intuitively satisfactory arithmetic if there is a counterexample to the appropriate U. The claim is that these considerations, suitably deployed, provide intuitionistic grounds for the affirmation of U which involve neither an inference from the presumed consistency of S nor any claim that a proof of (3) is somehow tantamount to a demonstration of U. Too good to be true?

In order to set matters up, we need to review a little of the general detail of Gödel's construction. As is familiar, the groundwork for his result is the assignment to each primitive symbol in the vocabulary of S, and thereby to each well-formed formula and sequence of such formulae, of a number — its *Gödel number.* This assignment is two-way effective; i.e., given an arbitrary formula, or sequence of formulae of S, we can in principle effectively compute its Gödel-number; and given an arbitrary number, we can effectively determine whether it is the Gödel-number either of a formula or a sequence of formulae of S and if so, to which formula or sequence of formulae it belongs. Crucially, as Gödel shows, such an assignment may be done in such a way that a certain primitive recursive relation, Pxy, holds between the numbers x and y just in case x is the Gödel-number of a sequence of formulae which constitutes a formal proof in S of a formula whose Gödel-number is y. That is, for all x and y,

Lemma 1 *Pxy iff. y is the Gödel-number of a wff, B, of S and x is the Gödel-number of a sequence of wffs of S which constitute a proof in S of B.*

The proof now proceeds by showing how the operation,

> substitution for each occurrence of the free variable with Gödel-number m in the formula whose Gödel-number is k by the numeral for n,

may be arithmetically represented; that is, how an effective function in three arguments, Sub $< k, m, n >$, may be arithmetically defined in such a way that, when k is the Gödel-number of a formula containing free occurrences of the variable with Gödel-number m, the value of Sub $< k, m, n >$ will precisely be the Gödel-number of the formula that results from substituting the numeral for n for each free occurrence of that variable. We then consider the open sentence

$$(\mathrm{U}^*)\quad (\forall x) \sim(Px, \mathrm{Sub} < y, m, y >),$$

where m is the Gödel-number assigned to the variable, 'y'. This will have a Gödel-number — say, g. So 'Sub $< g, m, g >$' will denote the Gödel-number of the formula that results from substitution for each occurrence of the free variable with Gödel-number m in the formula whose Gödel-number is g, viz. (U*), by the numeral for g; i.e., 'Sub $< g, m, g >$' will denote the Gödel-number of the formula

$$(\mathrm{U})\quad (\forall x) \sim(Px, \mathrm{Sub} < g, m, g >),$$

— the undecidable sentence itself.

Now recall the standard informal explanation of intuitionistic negation:

> **Negation** The negation of A is demonstrated by any construction which demonstrates that a contradiction could be demonstrated if we had a demonstration of A.[18]

The focus of what follows will be on the consequences of the supposition that an intuitively correct arithmetical system S contains a proof of '*Pkg*' for some particular choice of k. But to get any mileage, we now need to specify some assumptions about S. S is to be any formal arithmetic with a standard intuitionist logic which is strong enough to be within the scope of Gödel's theorems and such that:

> (1) S is *intuitionistically endorsed*, i.e. all the methods and assumptions of S are intuitionistically acceptable, so that every proof in S corresponds to an intuitionistic demonstration (I-demonstration). Note that this is *not* to assume the consistency of S; what is entailed is merely that if S is inconsistent, then so are the principles and assumptions incorporated in the informal notion of I-demonstrability;

and

> (2) S is *computationally adequate* — all arithmetical computations can be done in S. In particular, for each primitive recursive '*Axy*', and arbitrary 'k' and 'n', either '*Akn*' or '$\sim$*Akn*' is computationally decidable in S. Again, reflect that S can, of course, be inconsistent while having this feature.

(1) and (2) will be true of any formal arithmetic in which the intuitionist mathematician will be interested.

Now, suppose that

(i) '*Pkg*' is computationally verifiable.Then

(ii) '*Pkg*' is provable in S — by (2). But

(iii) U is provable in S — from (i) by Gödel's Lemma above. So

(iv) '$\sim$*Pkg*' is provable in S — since S has a standard intuitionist logic. So

[18]Thus Heyting in [85], p. 102: "$\neg P$ can be asserted if and only if we possess a construction which from the supposition that a construction [proving] P were carried out, leads to a contradiction. Compare Dummett in [47], p. 13: "A proof of $\neg$A is usually characterised as a construction of which we can recognise that, applied to any proof of A, it will yield a proof of a contradiction".

(v) '$Pkg \wedge \sim Pkg$' is provable in S — by (ii) and (iv).

It has thus been shown that, had we a computational verification of 'Pkg', we could accomplish a proof in S, and hence, by (1), an I-demonstration, of a contradiction. But any I-demonstration of 'Pkg' will be, in view of the primitive recursiveness of the predicate, a demonstration of its computational verifiability, i.e. of (i). By the foregoing, it will therefore constitute a demonstration of (v), so by (1) a demonstration that a contradiction is demonstrable. So, in the light of **Negation**, (i)–(v) constitutes an I-demonstration of '$\sim Pkg$'.

If that is right, it may seem that nothing obstructs advancing to a demonstration of U itself — indeed that, as Dummett in effect remarks,[19] the further step is merely trivial. Intuitionistically a demonstration of arithmetical '$(\forall x)(Ax)$' is any construction which we can recognise may be used, for an arbitrary natural number k, to accomplish a demonstration of 'Ak'.[20] (Call this principle *Generality*). Well, the reasoning from (i) to (v) will evidently go through, if at all, then for an arbitrary choice of 'k'. So it ought to be acknowledged to constitute, by *Generality*, an I-demonstration of U. QED.

It appears, then, that the reasoning from (i) to (v) constitutes, on the assumptions made, the basis of an intuitionistically cogent inference to U. That will constitute no advance on what we have already, of course, if those assumptions somehow smuggle in a presupposition of consistency, either of the system S or more generally, of the informal collection of principles and axioms which are sanctioned by the intuitionistic notion of demonstration. But, as noted, this does not appear to be so.

Why present this as distinctively intuitionistic demonstration — what happens if we try to transpose the reasoning to provide for a *classical* demonstration of U? The crux is the part played by the intuitionistic account of negation — that which licenses the claim that you have demonstrated '$\sim A$' as soon as you have shown how, given any demonstration of 'A', a demonstration could then be given that a contradiction was demonstrable. Naturally, this account will not be acceptable to the classicist in all cases; classically, truth is one thing and demonstrability quite another and there is no general reason why a classically *true* sentence should not merely be absolutely unprovable but such that if we could prove it, we could derive inconsistencies. However, might the in-

[19] See [39], p. 192.
[20] Cf. [85], p. 106, and [47], p. 12.

tuitionistic account be classically acceptable in the relevant case, where A is 'Pkg' and therefore effectively decidable?

Well, the difference the effective decidability of 'Pkg' makes is that either it or its negation is thereby guaranteed to be classically demonstrable. A proof that either cannot be demonstrated is therefore a proof that the other can be. Now, we have in (i) to (v) a classically acceptable proof that, had we a demonstration of 'Pkg', we could demonstrate the obtainability of a contradiction. *If* that amounted to a proof that no demonstration of 'Pkg' can be given, then a classical demonstration that each '$\sim Pkg$' is demonstrable, hence one of U, would be in prospect. But the problem is clear: no such conclusion follows unless we assume that it is classically impossible to demonstrate the classical obtainability of a contradiction. And that is tantamount to the forbidden assumption that the methods and principles incorporated in the classical notion of demonstration are consistent — a stronger assumption, indeed, than the antecedent of Putnam's conditional (3).

There appears, then, to be an asymmetry between the intuitionistic and classical cases. Intuitionistic demonstrations with no classical analogue are always interesting — more than interesting when the conclusion is arithmetical. Can it *really* be that we have one here?

There is one and, so far as I can see, only one serious doubt. It arises as soon as we reflect that the intuitionist ought not to be satisfied with *Generality* as formulated above. For that formulation puts no specific controls on demonstrations of 'Ak'. What is required is, rather, something along the lines:

> A demonstration of arithmetical '$(\forall x)(Ax)$' is any construction which we can recognise may be used, for an arbitrary numeral 'k', to accomplish a *constructive* demonstration of 'Ak'.

But what is a constructive demonstration in the case where 'Ak' is decidable by computation if not (a guarantee of the possibility of) the appropriate computation? Since '$\sim Pkg$' is such a case, the reasoning in the intuitionistic demonstration, while admittedly applicable to an arbitrary numeral 'k', ought to be reckoned as providing a basis for the universal introduction step to U only if it ensures that each '$\sim Pkg$' is *computable.* Does it?

Well, suppose, Heaven forfend!, that the wider notion of I-demonstrability *is* inconsistent, although the fragment consisting of arithmetical computation is consistent. Let S accordingly be inconsistent, and suppose 'Pkg' actually computationally verifiable. Then the informal demonstration of '$\sim Pkg$' still goes through, but now provides no assurance of that sentence's computational verifiability. Hence, to treat the demonstration as providing such an assurance is implicitly to assume that our supposition does not obtain — i.e. that I-demonstrability is consistent. Either, then, the universal introduction step is unjustified, since we have no assurance that each '$\sim Pkg$' is constructively verifiable; or, in taking it that we have such an assurance, we implicitly assume the consistency of I-demonstrability — and the forbidden assumption is at work in the demonstration after all.

This is a good objection as far as it goes. It rests, however, on the proposal that, where A is effectively decidable by computation, a constructive proof of A must consist in, or anyway establish the possibility of the appropriate computation. And other proposals are possible. Not all computationally decidable statements are atomic. So one possible proposal would be to grant that, in the atomic case, a constructive demonstration must consist in (proving the possibility of) the appropriate computation; but to allow, for certain non-atomic forms of computationally decidable statement, that certain kinds of proof besides the relevant kind of computation may count as constructive. In the crucial case, in particular, of the negations of atomic statements, it might be held to be sufficient for constructivity that (one prove that) *effective* means exist for arriving at a contradiction should a computational verification be provided of the statement negated. Since the decoding of 'Pkg' and consequent location of a proof in S of U, is an effective procedure, this more relaxed account would sanction the constructivity of the mode of demonstration outlined for each '$\sim Pkg$', and would — apparently without the need for the forbidden assumption — thus reinstate the intuitionistic demonstration of U under the aegis of the tightened version of *Generality*.

The status of the intuitionistic demonstration turns, then, on the vexed question of the proper interpretation of the notion of constructive proof — specifically on whether the intuitionist ought or ought not to allow that the more relaxed notion of constructivity outlined is what is relevant to '$\sim Pkg$'. But I cannot pursue the matter further here; clearly both proponent and opponent of the intuitionistic demonstration each have some considerable work to do.[21]

---

[21] By way of a couple of opening shots, here are two considerations whereby the

Let us take stock. I have now canvassed three ways whereby it might be contended that Gödel's theorem contributes to an informal demonstration, with respect to an appropriate system of arithmetic S, of the corresponding undecidable sentence U. But none has emerged as clearly successful. Unless some quite different approach has been overlooked, it would therefore appear that the widely accepted idea that, as Lucas expresses it, we are able to 'see' that U is true, has so far just not been made out.

Why *is* that idea so widely accepted? It certainly isn't because people accept the intuitionistic demonstration! I suspect the real reason is actually rather unflattering: that we succumb to a simple conflation, confusing the discovery of a *commitment* with the discovery of a *truth*. The proof of the conditional (3) for an intuitively correct S is a deeply impressive result which teaches us, on pain of accepting the inconsistency of our arithmetical thought, that we are *committed* to regarding each instance of U as computationally verifiable. Since we want to believe that our arithmetical thought is consistent — since, indeed, it is doubtful if agnosticism on the matter could be a practical option — there can therefore be no sitting on the fence as far as U is concerned. But it

---

proponent might at least begin to try to motivate the more relaxed notion of constructivity which the intuitionistic demonstration requires.

(i) Don't we need the more relaxed notion to make sense of the very problem case described? For that is a case where there is, by hypothesis, an I-demonstration — ergo a *constructive* demonstration — of U which can be given in S; and whatever else that is, it is required — by tightened *Generality* — to involve a construction which enables us to provide a *constructive* demonstration of '$\sim Pkg$', for arbitrary k. Yet the problem case is supposed to be one where no computational verification of '$\sim Pkg$' is possible. So either the problem case makes no intuitionistic sense, or the notion of constructive demonstration cannot be restricted in the way required to make a problem in the first place!

(ii) It will not in general be sensible to count proofs of truth-functional compounds of recursively decidable atomic statements as constructive only if they are constituted by the execution of the same class of recursive procedures. Such a proposal would restrict constructive proofs of, for instance, conditionals composed out of such statements to computational verifications of the consequent of falsifications of the antecedent — in no other case could such a conditional be regarded as constructively proved. But that would be to take all the conditionality out of the conditional, as it were — to abrogate the right to conditional claims in cases where one had no proof or disproof of either antecedent or consequent. The right account should allow such a conditional to have been constructively proved just in case we have shown how, given any computational verification of the antecedent, we could effectively get a computational verification of the consequent. If such a relaxation of the notion of constructive proof is appropriate for the conditional, why not for negation too?

is quite another thing to view this commitment to U as something we incur on the basis of a *demonstration* of its truth. We are, as a rough parallel, similarly committed by everything we ordinarily think and do to the existence of the material world; no agnosticism on the point is practical. It would be very much easier than it is to dispose of material world scepticism if this commitment could immediately be taken as the reflection of a cognitive achievement.

On this account, no informal demonstration of the undecidable sentence attends Gödel's proof, and nothing takes place worthy of signification as 'recognition' of its truth. We are merely brought to see that our standing commitment to the consistency of our arithmetical thought embraces a plethora of unsuspected, specifically arithmetical commitments, each associated with a Gödelian undecidable sentence. That is enormously interesting. But it provides no basis for claims about the indefinite extensibility of the concept of arithmetical proof, and no support whatever for Lucas and Penrose.

7. ON A NEW ARGUMENT THAT GÖDEL'S THEOREM SUPPORTS REALISM

Here is an appropriate point to make contact with a recent discussion of Christopher Peacocke's.[22] Peacocke argues that Gödel's construction does indeed subserve a simple, intuitively convincing train of thought which should be regarded as demonstrating the truth of U. But the argument, he contends, can go through only in the context of a *realist* semantics and is inaccessible on any constructivist (proof-conditional) conception of the meaning of quantification over the natural numbers. In Peacocke's view, then, Gödel's theorem, so far from underpinning anything like the specifically intuitionistic demonstration which we have just reviewed, presents instead a special problem for constructivism and is testimony to the superiority of semantical realism. However, his argument contrasts with the attempt, targeted by Dummett's discussion, to use Gödel's theorem to confound the thesis that meaning is given by use, a central plank of semantical anti-realism. It is important to Peacocke that the specific form of semantic realism which, he believes, is necessary for his simple informal demonstration is, in the form he favours, quite

[22] In Section 4, *Gödel's theorem: a problem for constructivism*, of his essay [125] in [77].

compatible with an insistence that meaning cannot transcend use but must be fully manifest in linguistic practice.

It is worth separating three main issues raised by Peacocke's discussion: first, the issue whether the argument he describes is, as he claims, properly taken as a demonstration that U, though unprovable in S, is true; second, whether the argument does indeed call for a realist understanding of the dominant quantifier in U; and third, whether the constructivist has no prospect of finding an analogue. I do not think Peacocke makes a convincing case for his view on any of these issues. I'll begin by briefly reviewing the third.

Peacocke reasonably insists that any constructivistic attempt at an informal demonstration of U must be sanctioned by an explicit account of the proof-conditions of universally quantified arithmetical statements. He proceeds to indicate various ways in which such an account might stumble. Suppose, for instance, we are offered the following (where 'A' is a quantifier free predicate):

> '$(\forall x)(Ax)$' is proved just in case we have a (constructive) proof that each statement of the form 'Ak', 'k' a numeral, is provable in primitive recursive arithmetic.

This is viciously circular — or, if not quite circular, close enough to circularity to be nugatory.[23] This complaint does not presuppose that any account falls into circularity merely by using the concept of universal quantification in an account of the proof-conditions of statements of the form '$(\forall x)(Ax)$'. No correct account could avoid circularity by that test. Rather, the point here concerns the *way* the explanans involves the concept of universal quantification: it *embeds* an occurrence of the universal quantifier within a that-clause serving to specify the content of the kind of proof it requires us to have if we are to count as having proved '$(\forall x)(Ax)$'. Since the totality quantified over, viz. statements of the form 'Ak', 'k' a numeral, is isomorphic to that of the natural numbers, the explanans presupposes a prior grasp of the proof-conditions of a species of universally quantified statement which is in no way interestingly different to the kind of statement whose proof-conditions it is supposed to explain.

That is one kind of abortive proposal. But of course it does not represent the constructivist's best shot. A better suggestion would run

[23] Peacocke does not use this actual example, preferring a (so it seemed to me, rather unnatural) formulation in which a strict circularity is involved.

along the lines of the revised version of *Generality* mooted towards the end of the preceding section:

> A demonstration of arithmetical '$(\forall x)(Ax)$' is any construction which we can recognise may be used, for an arbitrary numeral 'k', to accomplish a constructive demonstration of 'Ak'.

Here no grasp is presupposed, by the explanans, of what constitutes a proof of a universally quantified conclusion: the kinds of proof of which a grasp is presupposed are proofs of *singular* propositions of the form 'Ak'. Against this account — at least by the rules which Peacocke imposes on his own discussion — there is no legitimate complaint of circularity. What other complaint might be made?

Well, that nothing has so far been said about what, for the purposes of the explanans, should rank as a *constructive* demonstration. Peacocke writes:[24]

> We presume, when issues in the theory of meaning are not at stake, that what distinguishes a constructivist is the semantical clauses he accepts for various constructions. Our hold on the idea of constructivism begins to slip if these semantical clauses themselves contain the notion of constructive proof *ineliminably.* [My emphasis]

This is perfectly fair: the constructivist has to say (predicatively) what a constructive proof *is.* But Peacocke offers no clear reason for his apparent view that this is a demand which a constructivist will not be able to meet. And the fact is, to the contrary, that when our interest is in the intuitionistic demonstration already reviewed, there is no great problem about saying what a constructive proof is — no good reason to suppose that, *for the purposes of that specific demonstration*, an ineliminable play with the notion of constructive proof cannot be avoided. We reviewed two accounts of what should count as a constructive proof of '$\sim Pkg$'. On one, there was indeed a problem for the ambition of the intuitionistic demonstration to avoid any assumption of consistency. But on the other we found no such problem. Admittedly, the second made use of the arguably vague notion of an *effective* means for locating a contradiction; but there is no doubt that the means supplied by Gödel's own construction rank as effective under any reasonable meaning of the term. I conclude that Peacocke's claim, that no genuinely constructivis-

[24] See [125], p. 180.

tic demonstration of U is in prospect — that the constructivist who thinks he has such a demonstration is either neglecting his explanatory responsibilities or has surreptitiously gone realist — is overtaken by our discussion above.[25]

Now to the train of thought on which Peacocke bases his contention that Gödel's construction does subserve an informal demonstration of the undecidable sentence, but one which demands a classical understanding of universal quantification. Peacocke draws here on the 'commitment account' of universal quantification developed in his *Thoughts*.[26] For the purposes of *Proof and Truth* he offers the following simple outline:[27]

> [The commitment account] ... claims (a) that what makes it the case that someone is judging a content of the form 'All natural numbers are F' is that he thereby incurs this infinite family of commitments: to F0, to F1, to F2,... That someone has incurred this family rather than some other will be evidenced by the circumstances in which he is prepared to withdraw his judgement. The commitment account also claims (b) that a content of the form 'All natural numbers have property F' is true just in case all those commitments are fulfilled.

He adds:[28]

> The commitment account makes it relatively unproblematic that a first-order quantification should, though true, be unprovable from a particular recursive set of axioms. On the commitment account, what gives the universal quantification

---

[25] Peacocke frames his critical discussion around the candidate intuitionistic demonstration which he calls the 'Wright demonstration'. This is based on conversations and correspondence with the present author. But I have to warn the reader that, though similar in certain respects, this argument is *not* that with which we have been concerned. By contrast with the latter, Peacocke's 'Wright demonstration' features the soundness, and hence consistency of the object system as an explicit and undischarged assumption. Any such demonstration can have, of course, absolutely no advantage over a straightforward Dummettian modus ponens, and any play it makes with specifically intuitionistic understandings of negation and the universal quantifier will therefore be window dressing. However Peacocke's discussion is not weakened by its focus on the Wright demonstration, since his misgivings about it, as a purportedly constructive argument, are entirely focused on its final Universal Introduction step; and this it has in common with the more interesting intuitionistic demonstration with which we have been concerned.

[26] See [124], especially chapters 2 and 3.

[27] See [125], pp. 174–175.

[28] See [125], p. 175.

> its truth-condition is not the set of ways it can be proved, but the commitments incurred in judging it. The commitments of a first-order quantification may all be fulfilled, and hence the quantification may be true, even though it is not provable from the first-order axioms a subject is employing.

I have expressed reservations elsewhere[29] about Peacocke's claim that the commitment account is inherently *realist*, or makes realism 'relatively problematic'. It seems to me that someone could quite consistently accept the account — accept that the content of a universal quantification over natural numbers is individuated by the commitments incurred by someone who judges it true, and that such a quantification may be understood on just that basis by someone innocent of any concrete idea of what should count as a proof of it — yet still dispute that we have any conception of what it is for the infinite family of commitments associated with such a quantification to be fulfilled *undetectably*, in a fashion transcendent of proof. But it is important to see that this line of objection can be correct without subtracting all the interest of Peacocke's application of the commitment account in the present context. For, whether or not it implicitly involves a realist (epistemically unconstrained) idea of truth, the commitment account certainly is at odds with *constructivism*: if someone can distinctively manifest his acceptance and hence grasp of a universally quantified arithmetical content just by appropriate patterns of behaviour in relation to its (effectively decidable) commitments, then such understanding does not depend on grasp of proof-conditions, nor require any conception of what should count as a proof of such a content. If Peacocke's line of argument for the truth of U does, as he believes, depend on the commitment model, it will indeed be unavailable to *constructivists* even if not to all forms of *anti-realists* (verificationists).

The line of argument is very simple and runs as follows. Reflect that Gödel's argument ensures that if an arithmetical system is consistent, then for any natural number, k, the sentence '$\sim Pkg$' is provable in that system. And suppose that the particular system, S, we are concerned with is sound — has true axioms and truth-preserving rules of inference. In that case, Peacocke writes,[30]

> ...if for any k the sentence '$\sim Pkg$' is provable [in S], then for any k the sentence '$\sim Pkg$' is true. So all the commitments of

[29] See chapter 8, *A note on two realist lines of argument*, in [182].
[30] See [125], p. 178.

> the sentence '$(\forall x) \sim Pxg$' are fulfilled. But according to the second part of the commitment account,... a universal numerical quantification is true if and only if all its commitments are fulfilled. So on the commitment account, '$(\forall x) \sim Pxg$' is a true sentence.

However,

> The reasoning in this straightforward answer is unavailable to the constructivist. The reasoning relies on the commitment account which allows for the possibility, incompatible with any constructivism deserving of the name, that a universally quantified arithmetical sentence be true though unestablishable.

It should be evident, however, the reasoning *doesn't* properly rely on the commitment account — rather, as Peacocke himself remarks, it depends only on the second ingredient claim — claim (b) in the passage quoted. True, that claim refers to 'F0', 'F1', 'F2', etc., as 'commitments' of contents of the form, 'All natural numbers have property F'. But that is inessential: as far as its role in the reasoning is concerned, claim (b) comes to no more than — has no implications which are not shared by — the straightforward clause:

> 'All natural numbers have property F' is true if and only if for each number n, 'Fn' is true.

And rather than having any intrinsically anti-constructivist purport, this clause may readily be endorsed not just on non-epistemic conceptions of truth but on conceptions of truth, like that of the intuitionists as usually understood, whereby truth is held to consist in a kind of constructive provability and when the force of the biconditional is to claim that a constructive proof of either side is, or may be transformed into a constructive proof of the other. Since Gödel effectively establishes, by intuitionistically acceptable means, the intuitionistic provability of each '$\sim Pkg$', an intuitionist who accepts the straightforward clause, so interpreted, need have no difficulty in keeping company with Peacocke's simple demonstration.

It remains to observe, finally, that since Peacocke's 'demonstration' makes undischarged use of the assumption of the soundness, and hence consistency of the system of arithmetic concerned, its claim to *demonstrate* the undecidable sentence — to provide cogent a priori reason for supposing that sentence to be true — is no more impressive than that of the Dummettian modus ponens. (And since the conditional which provides the major premise for Dummettian modus ponens:

'If I is consistent, then U',

may be established by intuitionistically acceptable means for intuitionistically acceptable systems, I, Peacocke should never have been in doubt that an intuitionist who is prepared to take on the assumptions of Peacocke's own demonstration can construct a simple proof of the undecidable sentence.)

8. THE LUCAS/PENROSE ARGUMENT: CONCLUDING REFLECTIONS

Let me finish by offering some very summary suggestions about how matters should be regarded *if* a line of thought fit to be regarded as an informal demonstration of U, for an arbitrary intuitively acceptable arithmetical system, S, can after all be disclosed.

First on Lucas and Penrose. We have seen, in effect, that their argument has to be (i) that the concept of demonstration which governs human arithmetical practice is not the concept of demonstration which describes the operations of any Turing machine, and (ii) that the non-mechanical character of human mathematical thought is carried by this point. The claim will be that the former concept is shown, by the success of the hypothesised informal demonstration and the fact that such a demonstration is always available, to have an extension which admits of no effective enumeration. Given any effective enumeration (recursive axiomatization) of arithmetical truths, we have a method — contained in the technique for constructing the Gödel sentence and then proceeding, via the successful line of thought, to recognise its truth — for generating a new demonstration going beyond what can be accomplished, even in principle, by derivations from the axioms in question. In brief: the structure of the output-in-principle of the human mathematician and that of the output-in-principle of any Turing machine are different.

Two points seem to me worth logging concerning, respectively, the status of this disanalogy, and its capacity to carry an anti-mechanist conclusion. The first is a small but, in a context in which there has been much confusion about the role of suppositions of consistency, important qualification. Obviously, the disanalogy can be made out *only if* we take it that human arithmetical thought is consistent. Otherwise there is *of course* a Turing machine which generates all and only the arithmetical sentences of which we can in principle construct what, by our standards, rank as demonstrations. Even if the form of a demonstration of U can be disclosed in which the consistency of S does not feature as a premise, the claim to have shown thereby that, in general, the class of in principle humanly demonstrable arithmetical sentences is not effectively enumerable, will still depend on the assumption of the consistency of S and, indeed, of any intuitively acceptable arithmetic which strengthens S. The most that is in prospect, in other words, is still a *disjunctive* conclusion. But the disjunction is not the Gödelian disjunction cited in section 2 above. That disjunction featured as its second disjunct the proposition that the Turing machine which in fact encodes human arithmetical capacity is one whose formal specification no human being can comprehend. By contrast, the disjunction in prospect as the conclusion of the Lucas/Penrose line of thought supplants that disjunct by the (depressing) proposition that arithmetical demonstrability by arbitrary intuitively acceptable means is an inconsistent notion.

I said it was a small qualification. Clearly it is not a terribly damaging concession for Lucas and Penrose to have to make if their conclusion has to be not that:

> Human arithmetical thought is non-mechanical,

but that:

> Human arithmetical thought, if not inconsistent, is non-mechanical.

The latter would still be of considerable philosophical interest. There is, however — the second point — a question about the attainability of this conclusion on which our discussion has so far not impinged, and which seems to me very difficult. What we are assuming to be in prospect is a disanalogy, on the assumption of consistency, between the concept of demonstrability defined by the principles and methods which are intuitively acceptable to human mathematicians and any concept of demonstrability which governs the workings of a specific *Turing* machine. Since, as was stressed earlier, the feasible arithmetical output, so to speak, of even the most prodigious human mathematician can no doubt be matched and surpassed by a suitable Turing machine, making the disanalogy good will require reflection on the *intensions* of the relevant concepts of demonstrability. A sufficiently explicit characterisation will therefore be needed of the human notion, so to speak, to make it clear how, for any particular arithmetical Turing machine — still assuming consistency — an arithmetical demonstration lying beyond its scope might in principle effectively be found. Well, suppose that accomplished. Then the basis of the Lucas/Penrose thesis will have consisted in nothing other than the provision of an *effective* procedure for finding, for any particular recursive axiomatization of arithmetical demonstrations, an intuitively acceptable arithmetical demonstration not included within it.

Rather than striking a blow against mechanistic conceptions of the human intellect, there will therefore be an immediate question whether this whole trend of thought cannot at most disclose the inadequacy of the idea of a *Turing machine* as a stalking horse for mechanism. What an argument against mechanism ought to show is that, for the area of human thought where the mechanist thesis is contested, insight, imagination and creativity have a role to play which cannot be simulated by a mechanical model — which cannot be reduced to the implementation of any set of effective instructions. The great difficulty, always, is to render such ideas sufficiently precise to make them debatable, to make it clear what a defender or an opponent has to establish. But surely it is moot whether the debate as envisaged has succeeded in doing that. Whatever is shown by an argument which establishes that, for any particular recursive axiomatization of arithmetical truths, there exists — if human arithmetical thought is consistent — an *effective* procedure for constructing a demonstration of an arithmetical sentence not included in the list, it is *not* that human thought is essentially creative, gifted with a spark which transcends the merely mechanical implementation of any instructions which can be laid down in advance. Any consistent, recursively axiomatized system of arithmetic may be so specified that its Gödelisation *is* an effective procedure. And to the sentence which results from that procedure may then, as it were mindlessly, be applied whatever is the general form of the informal demonstration we are assuming has been provided.

True, the sentences which result from indefinite iteration of this procedure on an intuitively acceptable base arithmetic — say standard first-order Peano arithmetic — will not be recursively axiomatisable — will not coincide with the output of any particular Turing machine. But a proponent of Lucas and Penrose needs to say something to disarm the conservative response that a device or organism may be lacking that which opponents of mechanism wish to claim for the human mind even though there is no recursive enumeration of specifications of all the tasks which it is able to perform.

9. DUMMETT'S PROBLEM: CONCLUDING REFLECTIONS

The same basic point shows, I think, that — retaining the hypothesis of the demonstrability of U — there is no alternative but to view Michael Dummett's response to the problem about meaning and use as correct in its essentials. But I fear there may be some disagreement between Dummett and the present author about what the essentials are. The problem only arises on three premises:

(A) That the meaning of an expression has to be fully capturable, as it were, by some substantial *description* of its proper use;
(B) That in the case of the expressions 'sentence true of all the natural numbers' and 'ground for affirming that all natural numbers have a certain property', such a description will have to consist in or provide for an effective axiomatization;
(C) That any intuitively correct arithmetical system may be Gödelised and the appropriate U informally demonstrated to be true.

No-one is better aware than Dummett, of course, that much is unclear about premise (A). One would not expect, for instance, that it will always be possible substantially to characterise the proper use of an arbitrary expression *non-homophonically*; but homophonic characterisation cannot be the rule, or constructing specifications of correct use will become a triviality, and premise (B), that an axiomatic specification is wanted in the case of arithmetical truth, will be overturned too easily. Still, if (A) is to be granted, at least for the expressions 'sentence true of all the natural numbers' and 'ground for affirming that all natural numbers have a certain property', then, in whatever terms and format specifications of use ought generally to proceed, the solution to Dummett's problem must, on our hypothesis, consist in overturning (B). But now, if we are in position to affirm (C) — if the needed general line of demonstration has been made out — then the terms in which (B) ought to be overturned are surely clear. True, there is no complete axiomatic characterisation of the set of sentences which may be regarded as true of all the natural numbers, or of the set of acceptable arithmetical demonstrations. But we may lay it down as *part* of any characterisation that, for instance, where S is standard first-order arithmetic:

(i) Each axiom and theorem of S is true of all the natural numbers; and

(ii) Any sentence formed by applying to an intuitively acceptable, recursively axiomatised set of arithmetical truths the following procedure ... [and here we specify the construction of the sentence U for S] is to be regarded as true of all the natural numbers on the grounds ... [and here we apply to the sentence in question the general form of whatever we regard as the attendant informal demonstration].

This characterisation may very well not be exhaustive. But there is no doubt that it is non-trivial, does speak to the immediate issue raised by Gödel's theorem and, above all, is, as Dummett describes his own proposal, "as much in terms of *use* as any other"[31].

Since I claimed there is no alternative but to view Dummett's response to the problem as correct in essentials, it will be clear that what I am taking to be essential is just the thought that it is only in the light of a general characterisation of the contours of the informal demonstration that we can be entitled to claim that such a demonstration is *always* available. So the very justifiability of (C) presupposes that we already have, for the predicates in question, what (A) demands.

Now this simple line omits, of course, any mention either of the idea of indefinite extensibility, or, of the idea, which Dummett makes consequent upon it, that the notions, 'true of all the natural numbers' and 'ground for affirming that all natural numbers have a certain property', are *vague*. However, it seems to me questionable, in the light of Dummett's official formulations, whether, even assuming (C), Gödel's result is properly taken to show that the concepts in question *are* indefinitely extensible. And if that is right, we may here leave on one side the question whether indefinite extensibility is best viewed as a kind of vagueness. A typical passage is this:[32]

> ... It is precisely the concept of such a ground

— a ground for asserting that something is true of all the natural numbers —

> which is shown by Gödel's theorem to be indefinitely extensible; for any definite characterisation of a class of grounds for making an assertion about all natural numbers, there will be a natural

[31] See [39], p. 198.
[32] See [39], p. 194.

> extension of it. If we understand the word 'meaning'...so as to make the meaning of the expression 'natural number' involve not only the criterion for recognising a term as standing for a natural number, but also the criterion for asserting something about all natural numbers, then we have to recognise the meaning of 'natural number' as inherently vague.

Allowing that a demonstration of the undecidable sentence attends the Gödelian construction, it seems to me still not to follow that 'for *any* definite characterisation of a class of grounds for making an assertion about all natural numbers, there will be a natural extension of it'. The general *type* of ground for such an assertion, associated with the Gödel construction, will be definitely — or definitely enough — characterisable once and for all; for if it were not, there would be no basis for (C) and the problem would not arise. The 'natural extension' extends not *types* of ground, but particular *sets* of demonstrations captured by particular recursively axiomatised systems. What is always open to extension, that is to say, is not any definite characterisation of a class of grounds, but any recursive enumeration of a class of proofs. That is quite consistent with the availability of a once and for all characterisation of the particular type of ground which attends Gödel's construction and of the particular class of arithmetical truths thereby demonstrated.

To avoid misunderstanding, I am not claiming that it is possible to give an illuminating yet absolutely general characterisation of the extensions of the concept, 'true of all the natural numbers', or the concept, 'ground for affirming that all natural numbers have a certain property', nor ruling out that they may be vague in any respect. My point is merely the modest one that the situation generated by Gödel's theorem on the assumption of (C) is quite consistent with supposing that any characterisation should at least contain components corresponding to (i) and (ii) above. If that is right, many fascinating questions concerning the determinacy of the concept of natural number undoubtedly remain. But the immediate problem is, as it seems to me, disposed of.

University of St. Andrews