The Implications of Gödel's Theorem (*)

J.R. Lucas

Fellow of Merton College, Oxford
Fellow of the British Academy

I
In 1931 Kurt Gödel proved two theorems about the completeness and consistency of first-order arithmetic. Their implications for philosophy are profound. Many fashionable tenets are shown to be untenable: many traditional intuitions are vindicated by incontrovertible arguments. Gödel's original paper was long and difficult. Subsequent improvements have strengthened and streamlined the argument, but it remains detailed and technical. Still, although the proof is difficult to master, so that for most thinkers its validity has to be taken on trust, its general import is intelligible, and its implications easy to follow out.

The theorems apply to "First-Order Arithmetic", that is to Elementary Number Theory formulated in "First-Order Logic", in which we have:
1. the sentential connectives, negation, disjunction, conjunction, implication, (roughly equivalent to 'not', 'and/or', 'and', 'if') and the like,
2. the two quantifiers (Ax), for all x, and (Vx), (1) for some x, ranging over individual variables
3. a symbol, =, for identity.

It is a simple clear-cut logic, in which, given suitable axioms and rules of inference, we can prove as theorems everything that ought to be proved. Quine took it as the paradigm of logic. It is, to put it crudely, the sort of logic a computer could be programmed to do. The arithmetic the logic is used on is similarly simple. The only entities are the counting numbers, 1, 2, 3, 4, . . . etc., but not nought, the negative numbers, fractions, or the mathematicians' favourites, pi, e, \(\sqrt{-1}\) (the square root of minus one); and the only operations needed are addition and multiplication (though, it should be noted, both are required, and Gödel's theorems do not apply to restricted systems in which only one of these operations can be defined). To all intents and purposes the theorems apply to all mathematics: every reasonable mathematical system that can be expressed in first-order logic is, as far as that logic is concerned:
(1) incomplete, and
(2) not within that system provably consistent.

II
Gödel's great achievement was to produce a water-tight version of the Epimenides paradox. Epimenides, a Cretan, held that all Cretans always told lies, in which case, his statement must itself belie what was being asserted. More precisely, we can consider the statement 'This statement is false', and be led to an inconsistency either if we assume the statement to be true or if we assume it to be false. In time past philosophers evaded the paradox by arguing that the referring phrase 'This statement' failed actually to refer, but Gödel found a watertight way of referring to well-formed formulae expressing propositions by assigning to each a number. Properties of a well-formed formula (such as being an axiom) and relations between them (such as one well-formed formula following from another) could be represented as arithmetical properties of, and relations between, numbers respectively. Much of logic could thus be represented in terms of arithmetical properties of, and relations between, numbers, though, as Tarski had earlier discovered, the concept of truth could not be adequately represented in this way. Although unable to represent the concept of truth,
Gödel was able to represent that of "provability", "provability in whatever formal system had been chosen". So although unable to represent arithmetically the original Epimenides statement, he was able to work out an astronomically high number which would refer to the formula expressing the proposition
'This proposition is unprovable'
What exactly the number is will depend on the formal system and the numerical coding adopted. Rather than deal with any such number, which is, in these respects anyhow, arbitrary, I shall pretend that the number 1024 is the number that refers to the Gödelian formula. So we have:
1024 There is no proof within the system under consideration of formula no. 1024
If we consider this formula and work through each possibility in turn, we rule out the possibility of its being false and provable, false and unprovable, true and provable, and we are led to the conclusion that this formula is unprovable in the system but true none the less.

III
The immediate consequence is that truth cannot be defined in terms of provability. In any serious intellectual endeavour we shall be able to formulate simple mathematical arguments, and thus shall be subject to Gödel's incompleteness theorem. However far we go in formalising our canons of proof, we shall be able to devise propositions which are not, according to those canons, provable, but are none the less, true. So it is one thing to be provable, and a different thing to be true. Truth outruns provability. Many modern philosophers have held the opposite. The Logical Positivists of the Vienna Circle proposed the Verification Principle that the meaning of a proposition was its method of verification; Wittgenstein held that meaning was constituted by use, and use has often been construed as "assertibility conditions". But if truth outruns provability, then the meaning of a proposition claimed to be true cannot be given by its method of verification, and the use of terms is not confined to their assertibility conditions.

IV
The Gödelian formula cannot be proved. Hence its negation cannot be disproved. So I can assert its negation without being convicted of inconsistency. But its negation is false. I can assert a falsehood without its being proved to be false. It is possible to be wrong without being provably so. There are claims which cannot be proved or disproved, but which can be true and can be false. I can claim truth, but may be wrong. I am fallible. I make claims which may be false---but, also and importantly, may be true. I may have some warrant for them, but not a watertight guarantee; and though I may be subsequently vindicated and seen to have been right, I may also be found to be wrong and required to withdraw my original claim even though reasonable and well warranted when it was made. I often have to stick my neck out in the pursuit of truth. Nothing Venture, Nothing Win: if I want to have a chance of being right, I have to run the risk of being wrong.
If truth can outrun provability, reality can outrun knowledge. Many modern thinkers have thought to evade scepticism by cutting down reality to what can be securely known. Idealists, phenomenalists and mathematical intuitionists reconstrue our statements about reality as being really statements about knowledge, or about sense-data, or about mathematical proofs. But once we can separate the content of a truth-claim from the warrant for asserting it, we can allow that we may be making claims about reality even though we do not know them to be true. Warranted assertibility is one thing, intelligibility is another. I can understand Fermat's last theorem or Goldbach's conjecture even though I cannot prove them; and, similarly, speculations about quarks or the beginning of the universe. Verificationist arguments fail. But so also do all-embracing sceptical arguments. For although there is a difference between being provable in some specified system and being true, the truth of the Gödelian formula can be established incontrovertibly. There is no doubt that the Gödelian formula is true, and hence that the Gödelian formula is unprovable-in-the-given-system of first-order arithmetic, and hence that it is true. Thus it is provable, but not in the-given-system of first-order arithmetic. The sceptic is entitled to point out that the claims about reality go beyond the empirical evidence on which they are based, but not to maintain that there can be no other considerations leading us to affirm them. Although at any stage he may legitimately ask the
realist to prove his claims, he is not justified in laying down what form a proof must take or in
drawing widespread conclusions from the realist's failing to meet the challenge. In any system there
will be propositions we cannot prove, but it does not follow that they are not true—quite the
contrary.

V

Gödel's theorem has important consequences for the philosophy of mathematics, besides refuting
Intuitionism. Although not the only means of doing so, it provides a very simple proof that first-
order arithmetic admits of *non-standard* models. For since the Gödelian formula is unprovable-in-
the-given-system of first-order arithmetic, the conjunction of the axioms with the negation of the
Gödelian formula must be consistent—else we should have a proof of the Gödelian formula itself.
But if that conjunction is consistent, it must, by the completeness theorem of first-order logic, have
a model. And this model must differ from the standard model of the counting numbers together with
addition and multiplication, since it brings out the Gödelian formula as false whereas the in the
standard model the Gödelian formula is true. So we have a non-standard model of first-order
arithmetic. The axioms and rules of inference of first-order arithmetic fail to characterize it
adequately. Some forms of Formalism are thus shown to be untenable. Many mathematicians
embrace a Platonist alternative, arguing that they can recognize the standard model as correct and
the non-standard ones as defective, and that this is because they have direct acquaintance with the
counting numbers by some form of quasi-perceptual intuition, much as Plato—and Gödel himself—
averred. But this is not the only option. A Logicist alternative is to ignore Quine's strictures and
accept *second*-order logic—the logic in which we can quantify over predicates and relations as well
as individuals—as *pukka*. Once we are given second-order logic, we *can* characterize the
counting numbers adequately in the way Dedekind suggested. We characterize a first number 1 (or,
in some ways better, 0), and then say that the numbers are those that posses *all* the "hereditary"
properties possessed by the first number. The point at issue is best brought out by noting the
different formulations of Peano's Fifth Postulate, which in first-order logic takes the form:

\[ \{F(1) \& (Am)[F(m)\implies F(m+1)]\}\implies (An)F(n) \]

and in second-order logic takes the form:

\[ (AF)\{F(1) \& (Am)[F(m)\implies F(m+1)]\implies (An)F(n)\}, \]

the difference between them being that in the former F is a free variable, and in the latter is bound
by the universal quantifier (AF). In the former case we can replace F by any predicate we like,
whereas the latter must hold for *all* properties whatsoever, whether or not we like them, whether or
not we even know what they are. In the one case we can—and have to—pick the predicate referring
to the hereditary property holding of 1 and all its descendants, in the other a hand is waved over all
possible hereditary properties, and the counting numbers are just those that have all, but all, those
hereditary properties. Intuitively we can see that there could well be a difference, because if we are
to specify a predicate as in the former version of Peano's Fifth Postulate, it must be specifiable, and
there are only a denumerable infinity of specifiable predicates in any formal language, whereas we
let (AF) range over all properties, Hence it is only to be expected that while the more widely
ranging antecedent cuts the model down to size precisely, the more limited, and thus less stringent,
requirement of the first-order version of Peano's Fifth Postulate admits of other models too beside
the standard one.

When von Neumann heard of Gödel's theorem, he was lecturing on Hilbert's programme; he
immediately discontinued those lectures, and expounded Gödel's work instead, because Gödel's
work had conclusively shown Hilbert's Programme to be unattainable. Hilbert had hoped to rescue
mathematics from the paradoxes of Cantorian Set theory and Transfinite Arithmetic by
axiomatizing them, and then showing metamathematically that the axiomatic systems were
consistent. Granted a satisfactory consistency proof, a mathematician could continue working in a
theory, whatever objections the critics might raise, without fear of inconsistency. But Gödel's
Second Theorem showed that if first-order arithmetic is, indeed, consistent, it is impossible to prove
its consistency within first-order arithmetic itself. That is not to say that we cannot prove the
consistency of first-order arithmetic: Gentzen did so; but he had to use transfinite induction. And
the whole point of Hilbert's Programme was to secure the reliability of disputed branches of
mathematics by proving their consistency by methods so simple as to admit of no dispute. If
Cantor's Transfinite Arithmetic is in issue, it is no good proving consistency by means of transfinite
induction. Hilbert's hope of sicherheit is necessarily unattainable.

Quine was suspicious of second-order logic. Ontologically, it was, in his eyes, extravagant. If, as he
maintained, to be is to be the value of a variable, then second-order logic, which quantifies over
predicate variables (both monadic predicates, referring to properties, and polyadic predicates,
referring to relations) is committed, like Platonism, to the real existence of properties and relations.
Furthermore, second-order logic lacks various meta-logical features that were seen as merits of
first-order logic. The completeness and compactness theorems fail. Whereas in first-order logic
every formula that is true under all interpretations is a theorem, and can be proved from the axioms
by means of the rules of inference in a finite number of steps, there are in second-order logic many
well-formed formulae which are true under all reasonable interpretations but are not provable from
the axioms in a finite number of steps by means of the rules of inference. Second-order logic is not
"recursively axiomatizable": a computer cannot be programmed to produce from any given axioms,
by means of any properly specified rules of inference, all those formulae that are "logically valid",
that is, true under all reasonable interpretations. That second-order logic is in this sense not
mechanical is itself a consequence of Gödel's theorem, and shows that logic, if it be taken to include
second-order logic, is much more juicy than the Logical Positivists thought. Ayer dismissed the
whole of mathematics as just one big tautology. To make out that the propositions of logic are
merely analytic is a faceable claim so long as logic is taken to be first-order logic, for then anyone
who denied a logically valid proposition could be taken through aa proof of it step by step, and
forced at some stage into self-contradiction. But once we allow that logic includes second-order
logic, the charge of its being merely analytic loses its cogency. Not all logically valid well-formed
formulae---those true under all reasonable interpretations---are analytic; some are synthetic, though
not depending on experience for their truth, nor known only a posteriori. Synthetic a priori
propositions are no longer ruled out.

VI

If some logical truths can be synthetic, we may wonder, with Kant, how they can be known. Sense-
experience would be an unsuitable warrant, since it is characteristic of sense-experience that it can--
-logically can--be other than it is. Rather, we are led to affirm logical truths by rational
considerations, such as those adduced in the proof of Gödel's theorem, but ones that cannot be
completely codified in a set of canonical rules. Contrary to the teaching of many influential
philosophers in the Twentieth Century, being reasonable is not simply a matter of following a rule.
Reason outruns rules, just as truth outruns provability. Reason is creative and original. Although
any bit of reasoning can be codified into a set of rules, there will always be further exercises of
reason, however much we codify existing patterns of rationality, which go beyond the rules and yet
are evidently reasonable and right. Reason is a source of knowledge. It is not confined to empty
manipulations of experiential contents, or servile machinations towards passionate ends. Although
we do well also to attend to actual experience and be guided in our practical thinking by what
human beings actually want, we are not confined to those realms, but can assess their deliverances
rationally, and perhaps reach new conclusions which go beyond the empirical evidence or transcend
the motives and ambitions of contemporary men.

VII

Reason is creative and original. It goes beyond antecedently established canons of right reasoning.
And it can do so in a personal way, so that one man's original insight may differ from another's
without either being wrong. Gödel's theorem supports this view, because there is not just one
Gödelian sentence for each system, but many, depending on which scheme of numerical coding is
adopted. Gödel himself had a scheme based on the prime factors of each counting number, but other
schemes, less conceptually simple, but easier to manipulate mathematically, have been devised.
Even within the general schema adopted by Gödel, arbitrary permutations are possible—we could interchange the numbers assigned to & (and) and to V (and/or). It depends on the the coding precisely which numbers refer to axioms, and which arithmetical relation is that obtaining between a formal proof and the well-formed formula proved. Hence different codings will enable us to pick out different well-formed formulae as being themselves unprovable-in-the-system, and thus as being true though unprovable. But their truth is independent of the coding. So also is their unprovability-in-the-system. Truth is independent of all systems, and the coding is chosen only after the system has been specified. For any given formal system of first-order arithmetic there are many, many true arithmetical statements that cannot be proved in that system, but which might be picked out as expressed by the Gödelian formula if a suitable scheme of coding were adopted. Different individuals, adopting different perspectives, would be able to discern different true though unprovable propositions. And if this is so even within the austerely impersonal world of mathematics, we should not rule out the possibility of its being the case in the humanities likewise. We have been too long in thrall to a monolithic view of reason, supposing that it must yield just one right answer valid for all men in all places and at all times. And then we have felt that reason's uniform light obliterated all personal idiosyncrasy and individuality, and that real fulfilment was to be found in feeling and sensibility rather than rationality and common sense, and that the life of reason was a poor thing, cold and lacking all romance. But it is a false antithesis resting on a false view of reason. Reason not only can be original, but original in very variegated ways, well capable of accommodating the variety of individual genius.

VIII
Moral and political philosophy will be different once reason is allowed to regain its ancient sway. Instead of seeking—vainly as it has appeared—for some fundamental principle from which all else follows deductively, we can live with our actual, somewhat piecemeal, approach to moral reasoning and moral reform. Many different arguments, of many different types, may be legitimately adduced, depending on the actual question to be addressed. Although the way Gödel's theorem was proved follows a somewhat standard route, the upshot of the proof is that reasoning, even mathematical reasoning, comes in all sorts of novel forms, which we cannot anticipate but can recognise as cogent when it is presented to us. Solvitur ambulando: we do not need an exhaustive critique of pure reason, but can be content to hope that we shall continue to recognise good reasons when they are offered us. So, too, in politics, we can be much more relaxed once we are no longer compelled, on a priori grounds, to despair of reason. If reason is nothing, we should be pessimistic about the outcome of political debate, and compelled to view the political process as a crude conflict of particular interests, with the weaker always being trampled on by the strong, and with propaganda being the only way of bringing others to share one's views. Power-sharing would always, then, be a dangerous weakening of one's defences, and indoctrination the only means of transmitting one's values to future generations. Many élites have excluded all others, fearing to let go of the ears of the wolf, and often the bitterest quarrels have been over the control of education. But once reason is acknowledged to have some sway over men's minds, the case is greatly altered. We can do business with reasonable men, knowing that should we concede the force of their arguments they will not automatically construe our concession as a sign of weakness, but will be the readier in their turn to grant the cogency of good arguments adduced by us. And, while always aware of the importance of education, our attitude will be very different if in the end we rely on the unfettered judgement of educated men rather than the conditioned responses of those whose upbringing we have been able to control. If the Other Side is Wrong, their arguments, however meretricious, will not in the end survive exposure to ours. Magna est Veritas, we shall be justified in believing, et Praevalebit.

IX
The thrust of Gödel's theorem is revolutionary, but in a conservative direction. Many old truths, discarded by modern fashionable philosophers, are rehabilitated; in some cases shown to be definitely true, in others not to be necessarily false. Claims which would rule out traditional tenets and natural patterns of reasoning can be articulated precisely and formally in mathematical logic,
and then shown not to hold in one central, incontrovertible instance. The instance is the Achilles'
heel of all wide-ranging metaphysical systems, that is the consideration of how they fare when
assessed according to the precepts they promulgate. Many in that trial are hoist with their own
petard. In general, however, it is difficult to get a grip on the argument, because it is inherently self-
referential, and self-reference is tricky, and subject to reasonable doubts as to whether the referring
expression actually refers. Gödel's theorem overcomes this difficulty. It overcomes the normal
slipperiness of wide-ranging metaphysical arguments by concentrating on a small fragment of
mathematics formalised in precise logical terms. A complicated coding system guarantees
reference. In most cases that would still be not enough to secure self-reference, but Gödel was
dealing with an infinite set of numbers, and it is one of the peculiar features---a defining feature
according to Dedekind's definition of infinity---of infinite sets that they can be mapped by a one-
one correlation into a proper subset of themselves. (3) So it is possible for an infinite set to contain
within itself a complete model of itself, in a way that is not possible for a finite model (e.g. a model
village, containing a model of the model village, itself containing a model of the model village,
that containing . . .). Once we embrace infinity we open the way to a coding that can code the
whole system and yet still be part of it. Infinity---Dedekindian infinity---makes Gödelian self-
reference possible, and thus an incontrovertible instance in which limited and finitely specified
accounts of truth and rationality are evidently inadequate.

Notes

(*) A talk given in Manchester in November 1996. For the copyright of the papers of J.R. Lucas see
http://users.ox.ac.uk/~jrlucas/ back
(1) For ease of typewriting, I use capital A, instead of the inverted A, to signify the universal
quantifier, and capital V, instead of the backwards E, to signify the existential quantifier; besides
ease of typewriting, these symbols have the merit in being in line with the standard signs for
conjunction and disjunction, for intersection and union, and with caps and cups. back
(2) For many illuminating examples, see Roger Penrose "On Understanding Understanding"
forthcoming in International Studies in Philosophy of Science, January 1997. back
(3) For example, if we correlate to each number the number that results from multiplying it by 2, we
have a mapping of all the counting numbers
1, 2, 3, 4, 5, . . .
into just the even numbers
2, 4, 6, 8,.10,
which seems to suggest that there are just as many counting numbers as there are even numbers,
although the even numbers are evidently only some of the counting numbers, since the counting
numbers include also the odd numbers, 1, 3, 5, which are not in the set of the even numbers. This
paradox greatly puzzled thinkers in late antiquity and the middle ages, and was only resolved in the
Nineteenth Century by distinguishing the equality generated by one-one correlations from that
generated by set-inclusion. back

from David Miller, University of Warwick. Date: Thu, 14 May 1998 13:30:10 +0100 (BST)
Dear John
Almost by chance I came across your two talks on the Web. Since they are clearly designed for a
general audience, and since I teach this term a short course on the theme The Liar to Gödel I am
moved to make a few critical comments, not on the principal thesis but on the details. I hope that
you do not mind.
in I You do not mention ' (successor) or 0 in addition to the operations of + and x, and I wonder if
that is deliberate. What you say here is not exactly wrong, but it is a little misleading. By
Presburger's theorem arithmetic with + and ' is decidable, and Skolem showed that arithmetic with x
alone is decidable. If ' is added, addition becomes definable by \( i + j = k \text{ iff } (i' x k'')' x (j' x k'')' = ((i' x j')' x (k'' x k''))' \) (Boolos & Jeffrey, 3rd edition, p.219). Presumably we can define 1 in arithmetic with + and x by \( k = 1 \text{ iff } (Ai) i x k = i \), and hence define \( i' = i + 1 \). It is not, however, quite correct to say that Gödel's theorems apply to 'every reasonable mathematical system'. They do not apply to geometry, whose decidability was proved by Tarski.

In VII You are right, of course, that the exact import of the Gödel sentence may depend on the system of Gödel numbering. But given that all useful systems of numbering are primitive recursive, and hence primitive-recursively intertranslatable, does it not follow that all such Gödel sentences are actually equivalent within Peano arithmetic? There are, of course other undecidable sentences that are not equivalent to any sentence of the form "I am not a theorem".

Kind regards - David Miller

Dear David,

Thank you very much for your E-mail, which I found waiting for me on my return last night. It is always very valuable to have feedback and criticism, and I am grateful to you for taking the trouble to query dubious points.

The omission of Presburger and Skolem was deliberate. It is quite surprising that arithmetic with + alone, and arithmetic with X alone, should be decidable, but it would only complicate the account without (as far as I can see) illuminating the main point. For the same reason I did not deal with successor: almost everyone is familiar with + and X, but only one in a thousand has heard of successor.

The omission of Tarski and geometry was not deliberate, but an oversight. I ought to have chosen a phrase which covered most of mathematics, but did not claim the lot.

I will print at the end of this a passage from a book on the foundations of mathematics, which I hope will forthcoming soon.

In your comment on VII I am not clear what you mean by 'equivalent'? All I wanted to get across is that there is not just one, rather recherché, wff that is undecidable in a given system, but lots. They may be equivalent in some sense, but they will be making different arithmetical claims about different numbers.

Thank you v. much for writing.

Best wishes

John

From: David Miller, University of Warwick Date: Mon, 18 May 1998 12:25:05 +0100 (BST) We largely agree. That's good. One or two rejoinders.

You write:

===> `VII I am not clear what you mean by 'equivalent'? All I wanted to get across is that there is not just one, rather recherché; wff that is undecidable in a given system, but lots. They may be equivalent in some sense, but they will be making different arithmetical claims about different numbers."

I mean equivalent within PA. Of course there are infinitely many undecidable sentences of PA no one of which is equivalent (within PA) to any other. This is not anything to do with the different ways of coding. For fix some system of Gödel numbering. Let G be the (unique) sentence that the diagonal lemma produces for PA; G says "G is not a theorem of PA". Since G is undecidable in PA we can now apply the diagonal lemma to PA + G. This produces a new Gödel sentence G*. And so on. Clearly, G and G* (and G** and G*** and ...) are not equivalent, either in PA or in PA + G. (G* is not a theorem of PA + G, plainly; a fortiori it is not a theorem of PA alone.) This has nothing to do with the use of different systems of coding, and all to do with the fact that PA is essentially undecidable [no consistent extension is decidable].

In other words, I quite agree with your conclusion, but I think that you attribute it to the wrong cause.
All good wishes - David

Dear David,
We both agree that there are lots of undecidable sentences in PA. I still maintain that this can be established by reference to different coding schemes---but concede that your argument is a better one (deeper; more illuminating). In my Minds, Machines and Gödel I used the infinite sequence G and G* and G** and G*** and ... to fox the mechanist, and should have used them to make the general point of there being an infinite number of undecidable sentences in any system of PA.
Your criticisms of my more technical paper I will incorporate there.
Again, many thanks
John

from ch.8, section 6, of The Conceptual Roots of Mathematics, which I hope will see the light of day soon.

Rather surprisingly, neither addition by itself nor multiplication by itself is enough. Presburger has shown that if we have only addition but not multiplication, the theory is complete and decidable. Similarly Skolem showed that if we have only multiplication but not addition, the theory is, again, complete and decidable. Even more surprisingly Tarski has shown that geometry and the elementary algebra of real numbers, with addition and multiplication, but without the general notion of natural number, is complete and decidable.¹

These results are surprising, and call for some explanation. The reason why Gödel's theorem does not apply is that in these theories we are thinking of the numbers in some other way---e.g. as an infinite Abelian group with addition as the composition function: they are not counting numbers, an endless succession of individually distinct entities, capable of being used for coding without ever being in any danger of running out. Once we have the natural numbers as counting numbers, we are subject to Gödel's theorem, because there is a potential infinity of natural numbers, and Dedekind-infinite sets can be correlated with a proper subset of themselves. But we need only a potential, not an actual, infinity of natural numbers. Gödel's theorem holds not only of Peano Arithmetic, but of the simpler Sorites Arithmetic. Essentially, that is, we can address Gödel's argument to a computer. We do not need to invoke the hand-waving arguments of Chapter 6 to avail ourselves of the principle of recursive reasoning, in order to prove Gödel's theorem: it applies even within the austere limitations of Sorites Arithmetic, with no commitment to any actual infinity.


last revised May 23, 1998