S & CB (1991) 3, 51–55 0954–4194

NIGEL J. CUTLAND

What does Gödel tell us?

A response to Dr. Cundy's article.

Key Words: Apologetics, Artificial Intelligence, Church–Turing Thesis, Gödel, proof, truth.

This is the age of computers and information technology, so it is not surprising that the mathematical results of Gödel and his contemporaries Church, Turing and Tarski in the 1930's continue to arouse interest. The fundamental limitations of formal mathematical systems and computer programs that they discovered then are directly relevant to computer science and information technology today. Moreover, as Dr. Cundy points out, these results are claimed to be applicable outside the field of mathematics and computing; it has been argued, for example, that Gödel's theorem shows that man's mind cannot be modelled by a computer program, and that artificial intelligence (AI) is not possible. Clearly such an application is relevant for Christian thought and apologetics, so the question of the validity of this and other purported applications of Gödel's theorem is important.

Gödel's first theorem (as refined by Rosser) applies to any system S of axioms and rules of proof that includes (implicitly or explicitly) the basic axioms of arithmetic. The Gödel sentence G (which can be constructed explicitly provided we know S) has the property that, if we are given a proof of G (using the rules of S) then that proof can be converted into a proof of not-G; and vice-versa. So, if S is consistent (i.e. free from contradiction), then neither G nor not-G can be proved using the rules of S. S is *incomplete*.

It is worth noting that there is no reference here to the *truth* of G, and we mention below the problem of defining truth; however, on any *informal* understanding of the classical notion of truth, we should say that either G or not-G is true. (Here it turns out to be G, because there is a standard listing of all statements, in which G appears as the m^{th} statement, and G has been constructed to have the informal meaning 'the m^{th} statement is unprovable'.)

An equivalent computer version of Gödel's first theorem is this: suppose that a computer C is systematically printing out statements of arithmetic in a consistent way, and it has the basic rules of logic and axioms of arithmetic built into its program. Then neither the statement G nor the statement not-G will ever be printed out by C.

Artificial Intelligence

The argument advanced by Lucas,[1] and endorsed by others, that uses Gödel's first theorem to show that a man's mind is more than a computer, and hence AI is not possible, seems to run as follows: An intelligent person, if he had the capabilities of a machine C and was able consistently to generate and print out statements of arithmetic that he knows to be true, would be able to do Gödel's construction of the true statement G, and hence output G—the statement that the machine C will *never* print out. So he is more intelligent than C—and his mind cannot be merely a sophisticated machine. He can always keep one step ahead of (or 'out-Gödel') any computer. This argument has also been used to refute one modern version of the belief that the universe is determined—namely, that it is no more than one gigantic computer (begging the question, of course, as to who wrote the program).

As Dr. Cundy points out, not all (himself included) are convinced of what I will call the *Lucas argument*. Of course, if it is *invalid*, then this does not prove that AI is possible, still less that the universe is determined, even though some who disagree with Lucas (such as Hofstadter) seem to do so from the position of belief in the possibility of AI.

Personally, I am not convinced by the Lucas argument. It seems to overlook the fact that in order to produce the Gödel statement G it is necessary to have a complete specification of the computer C and its program. So, personalizing the Lucas argument, if I am to out-Gödel the machine C that I am supposed to be, I have to know in principle everything about my current configuration and the program that controls me. I am inclined to think that this is too great an assumption to carry the argument through. It is a matter of speculation as to whether such total self-knowledge is possible. If it were, then Lucas' argument might have some force. It may be that Lucas is imagining someone else having complete knowledge of me *and telling me*: '. . . the mind . . . can pick a hole in any formal system presented to it as a model of its own workings.'[2] But to tell me my own specification is to *change* my specification—and the argument breaks down. Of course, to allow inputs from outside is entirely realistic, so any mechanistic hypothesis has to apply to the totality of minds. Then, for the Lucas argument to be valid, it would be necessary to assume the possibility of complete knowledge of *all* minds. Similarly, to use the Lucas argument to demonstrate that the universe is not a gigantic computer, it would be necessary to assume our ability to know its program completely if it were. Only then could we obtain the Gödel statement G showing that it is not, after all, such a giant machine.

Before leaving the discussion of the bearing of Gödel on AI, I would point out what I believe to be an error of understanding by Dr. Cundy regarding the *Church–Turing Thesis*. Historically this is the thesis that the

1 Lucas, J. R. Minds, Machines and Gödel, in A. R. Anderson (ed), *Minds and Machines*.

2 Lucas, J. R. op. cit. p. 49.

undefined, informally understood idea of mechanical computability is captured precisely by the notion of a Turing machine or, equivalently, by any modern digital computer with unlimited memory. Dr. Cundy seems to equate this restricted thesis (which is accepted by most working mathematicians) with the belief that all mental processes can be simulated by a Turing machine or digital computer. This is a much more extravagant claim, and is called the *Church–Turing Thesis—AI Version* by Hofstadter.[3] It is essentially the thesis that minds are nothing but machines, and goes well beyond what Church or Turing were suggesting, in my view.

Christian Apologetics

Returning to Gödel's work, let me mention briefly some ways in which I believe it *can* help in Christian apologetics. The first is rather simple, and is related to the argument of Jaki[4] referred to by Dr. Cundy. He addresses the question 'Is there a unified theory of the universe that is necessarily true?'. Any such theory would presumably be a consistent extension of the theory of arithmetic, obeying the usual laws of logic, so Gödel's first theorem will give us a statement G such that neither G nor not-G is settled by the theory. So there can be no *complete* unified theory of the universe; and from the purely logical point of view it is consistent either to add G or to add not-G to the theory. If either of these statements were *necessarily* true, then we would already find them in the theory. Somehow, if we were able to get outside the theory we might feel strongly that G is true (for reasons already explained), but presumably the nature of a grand unified theory is that you cannot get outside it. Thinking now in terms of any systematic philosophy or theology—stated in ordinary finite language—then we could argue that this will also be incomplete, and fail to account for some things that are so evidently true (like G) that we might even say they are necessarily true. (Of course, all this ties up with what we as Christians believe anyway—being finite, we cannot know all about God.)

A second application that I believe is helpful uses Gödel's second theorem—which shows that *if* a given system extending arithmetic is consistent (i.e. free from inner logical contradiction), then this cannot be *proved* without appeal to some higher principle—such as transfinite induction—from *outside* the system. But this begs the question of the reliability of this higher principle (is *it* consistent?). And in fact this argument applies to the whole enterprise of mathematics: if that is consistent, we have no way to prove it without appeal to some principle *outside of mathematics*. As working mathematicians we *believe* our subject to be free from contradiction, but this has to be ultimately a matter of faith. Eves and Newsom[5] express it thus:

'Suppose we loosely define a religion as any discipline whose

3 Hofstadter, D. R. *Gödel, Escher, Bach: an Eternal Golden Braid*, Basic Books, New York (1979).
4 Jaki, S. L. Review of Hawking, *A Brief History of Time*, in *Reflections* vol 7 (1988).
5 Eves, H. & Newsom, C. *An Introduction to the Foundations and Fundamental Concepts of Mathematics*, Holt, Rinehart & Winston (1966) p. 305.

foundations rest on an element of faith ... [then] mathematics would hold the unique position of being the only branch of theology possessing a rigorous demonstration of the fact that it should be so classified.'

Neuhouser,[6] quoting Eves and Newsom then applies Gödel's second theorem to demonstrate the *necessity* of faith when it comes to anything more advanced than arithmetic, and thus we can 'prove' that it is not *unreasonable* to hold the Christian faith. Of course, to be reasonable, this must be based on evidence (1 Peter 3:15), just as our belief in the consistency of mathematics is not blind faith, but rests on cumulative evidence. (There is a connection here with the Church–Turing thesis mentioned above, which is *believed* (on the basis of evidence) by most working mathematicians, but by definition cannot be *proved*, because it equates an informal idea with a precisely defined mathematical idea.)

Truth and Provability

Gödel's first theorem can be seen as pointing up the distinction between *truth* and *provability*. Provability *can* be defined in the framework of arithmetic (that is what makes Gödel's theorems work), whereas an important theorem of Tarski showed that the idea of a true statement of arithmetic *cannot* be defined in this framework. In the larger arena of mathematics as a whole, the meaning of mathematical truth is a fundamental philosophical question that mathematics itself (by Tarski's theorem) cannot answer, so again appeal must be made outside the system. Perhaps the work of Gödel and Tarski here may validate the distinction between *truth* and *truth that we can know*. We may be inclined to think of the latter as beginning with propositional truth; but then Jesus' statement 'I am the Truth' suggests that truth is much more than this. He did not merely claim to make true statements. Truth is ultimately personal, and Tarski's result showing the need to go outside of any formal language to define truth in it shows that this is a consistent position to hold. Certainly Gödel's first incompleteness theorem, together with Church's theorem (1936) on the inability of any computer program to decide whether a given statement is provable (let alone true) indicates a severe restriction on the nature and extent of what we can know—at least without some kind of external input or revelation.

Again, the work of Gödel and his contemporaries can be seen as giving a clear demonstration of the limitations of finite language. This is helpful evidence in support of the scriptural view that language and communication is much more than words in languages of symbols: Jesus is *the* Word; God has spoken to us *in* Son (Heb. 1:2)—not in Greek or English. Ultimately, it seems, language, like truth, is embodied in a person; all other notions of

6 Neuhouser, D. L. Truth: Mathematical and Biblical, *Journal of the American Scientific Affiliation* (1979) 31 pp. 29–33.

language and truth have inherent limitations—and Gödel and Tarski corroborate this view.

Dr. Cundy rightly endorses the need for great care in applying Gödel's theorems in disciplines outside mathematics and computer science. Certainly it would be nonsense to try to treat the Bible as a formal system; taken even at its very lowest level it is like a number of two-dimensional pictures of a higher dimensional object—each consistent in itself but together perhaps *apparently* inconsistent—or paradoxical. But in mathematics paradoxes can be resolved by moving into higher dimensions, or taking more care with the meaning of words and the use of logic; and similarly with the Bible, describing as it does an infinite personality. This raises another reason for caution in applying Gödel's results to matters of philosophy or theology. His theorems employed classical logic, extrapolated from our experience in the *finite* world. There is no necessary reason to assume that this or any other particular logic is appropriate to ultimate or infinite things—and the same comment can be made of the classical notion of truth. We know, for example, that infinite numbers behave differently from finite numbers: if m is an infinite cardinal number, then $m + m = m = m \times m$, and other strange things happen. And if, in mathematics, we have to be careful when reasoning about the infinite, even more so in matters of theology or philosophy.

Nigel J. Cutland is Professor of Pure Mathematics at the University of Hull.