Etica & Politica / Ethics & Politics, 2003, 1
http://www.units.it/~dipfilo/etica_e_politica/2003_1/5_monographica.htm

Human and Machine Logic (*)

I.J. Good

Trinity College, Oxford and Science Research Council, Chilton

The following summarising paragraph contains a few terms not defined until later:
Given any consistent formal system containing arithmetic, a man who has understood Gödel's construction can write down a true theorem G expressible in the system but unprovable in it. (The man will believe G if he believes the system is consistent.) A machine program that represents the formal system will never print G, even if the program contains a randomising device enabling it to apply the rules of inference in an arbitrary order. It has been argued that this shows that a man, *qua* mathematician, transcends a machine in at least one respect. (The argument does not depend on whether machines are capable of belief, nor on whether they could act as if they were so capable. I think they could act (1) so but the reader need not worry about this point since the present note is concerned with logic, not with probability.) This point of view is essentially refuted by the observation that Gödel's construction could itself be carried out by another (deterministic) machine. But further Gödel propositions can then be appended and a complete treatment leads inevitably to questions concerning the formalisation of transfinite counting, which incidentally preceded Gödel's construction historically by several decades. If the mentalists still wish to make a case they must base it on transfinite counting rather than on Gödel's theorem. It is entertaining to note that transfinite counting can be vividly expressed in polytheological terms. So much by way of introduction.
Lucas (2) argued that 'Gödel's theorem seems to me to prove that Mechanism is false, that is, that minds cannot be explained as machines'. Feeling that mathematical results can sometimes be proved by metaphysical arguments, but not conversely, I argued (3) that there must be some loophole in Lucas's thesis. But my discussion itself contained an error so I am anxious to argue the case again with greater accuracy. This accuracy is bought at the cost of an increase in technicality, but I believe the arguments will be intelligible at least to all philosophers of science.
Given any computer program (formal system) for proving theorems in arithmetic, Gödel's construction (4) enables us to print new theorems which the original program would not print, and the new theorems are true if ordinary arithmetic is consistent. Let us express that more carefully.
A (*finitely based*) *formal system* is defined in the following manner. We are given a finite alphabet of symbols, and a finite set A of finite strings of these symbols, each string being called an *axiom*. Axioms are regarded as a special case of *theorems*. We are also given a finite set R of *rules of inference* which can operate on some finite sequences of theorems, each such operation produces a new theorem (again a finite string of symbols) and this theorem is then said to be *proved*. We call this the formal system F = (R, A).
Corresponding to any such system, a computer program can be written which will print in turn each theorem of the system. The number of theorems is usually infinite, but *each* theorem (if provable) will ultimately be proved (printed).
The Gödel construction C can be applied to any finitely based formal system F, provided that ordinary arithmetic is represented in the system, and will yield a new 'theorem' or 'proposition' (string of symbols) G, called a Gödel proposition. This proposition will denote an

arithmetical statement that is true in the formal system F, but is unprovable provided that F is *consistent*. (In an *inconsistent* system every proposition is provable including the 'false' ones such as 0 = I.) Moreover, if a system F_1 is consistent, its Gödel proposition G_1 can be appended to its set A_1 of axioms, and the new system F_2 will still be consistent.

In order to avoid a proliferation of notation, let us denote by F a program that prints the theorems that the system F can prove. Now the Gödel construction C can itself be expressed as a program, which we also denote by C. When C operates on the program F_1, it produces a program F_2, this being a representation of the formal system F_2. Since the program F_1 *is itself a string of symbols* we begin to regard the program C as itself a formal system, but since it can be applied to every program of the form (R, A) it might be very difficult to show that it is finitely based.

The new formal system F_2 satisfies the requirements for the application of Gödel's construction. This will give rise to a new Gödel proposition G_2 and a new formal system F_3 and so on.

We can imagine a human operator playing a game of one-upmanship against a programmed computer. If the program is F_n, the human operator can print the theorem G_n, which the programmed computer, or, if you prefer, the program, would never print, if it is consistent. This is true for each whole number n, but the victory is a hollow one since a second computer, loaded with the program C, could put the human operator out of job. And even the original computer, suitably programmed, would be able to print in turn each of the Gödel propositions G_1, G_2, G_3, … , by repeatedly applying the operator C. By means of a process known to logicians as *triangularisation*, this program could be modified to print each of the theorems of the infinite sequence of formal systems F_1, F_2, F_3 … It is natural to denote this program by F_ω, where ω is the first *transfinite ordinal*. In spite of appearances, F_ω could be written in finite terms and correspond to a finitely based formal system (R_ω, A) where A is the same finite set of axioms that we started with, that is, the axioms of the system F, and R_ω is R plus a finite number of extra rules.

The notion of transfinite ordinals can be thought of in terms of polytheism. We imagine that, for each integer *n*, $ZEUS_{n+1}$ made $ZEUS_n$, where $ZEUS_\omega$ made $ZEUS_n$ for all *n*. Who made $ZEUS_\omega$? Answer: $ZEUS_{\omega+1}$, and the suffixes can be continued indefinitely, thus: I, 2, 3, … , ω, $\omega+1$, $\omega+2$, … , 2ω, $2\omega+1$, … 3ω, … , ω^2, … , ω^3, ..., ω^ω, … , ω^{ω^ω}, ….

Similarly, the Gödel construction C can be applied to F_ω, giving $F_{\omega+1}$, and we can proceed to higher and higher systems, just as in the process of transfinite counting, sometimes adding I, and sometimes applying a generalisation of triangularisation.

In order to write a program that can carry out this construction as far as any specifiable ordinal, it will be necessary at least to invent a representation for this, and for all smaller ordinals, on the integers. There is a known complete process for doing this, and the process of transfinite counting is thus naturally described as 'creative'. The use of this term is not evidence that the process cannot be formalised, and I believe that a sufficiently well-written program would be able to go as far in transfinite counting as any man can ever go. It is useless for the 'mentalist' to argue that any given program can always be improved, since the process for improving programs can presumably be programmed also; certainly this can be done if the mentalist describes how the improvement is to be made. If he does not give such a description, then ha has not made a case.

A similar controversy applies in a wider context: the only reason I know to suppose that the creative intellectual process of man cannot be mechanised is the weak one that it has not yet been done. I am of course here ignoring such practicalities as cost.

If the controversial 'axiom of choice' is true, then there is a smallest unconstructible transfinite ordinal τ. The question of its 'existence' is somewhat controversial, but of course it cannot be

reached by any transfinite counting program. The controversy is bound up with what is meant by mathematical existence. ZEUS$_\tau$ should have a prominent place in any polytheology.

Some readers will have asked themselves what meaning it can have to say that a proposition, expressible in a formal system, is 'true' if it is not provable. One answer is that a proposition P, of finite length, can express an infinite number of provable propositions, and yet perhaps not itself be provable, since a proof, by definition, must be finite. (5) An example of a proposition that might be of this form is 'For all positive integers, *r*, *s*, *t*, and *n*, we have $r^{n+2}+s^{n+2}\neq t^{n+2}$'. This is of course the famous unproved 'Fermat's Last Theorem'. Like Riemann's hypothesis concerning the zeros of the zeta function, this 'theorem' might be true but unprovable but if false it is provably so.

An error in my *New Scientist* article, which was pointed out by Alan L. Tritter (who has also made many other useful suggestions), was in the assumption that the finiteness of the internal storage of a computer (or man) would prevent it (or Him) from attaining some constructible infinite ordinals. The limitation cannot be in the finiteness of the internal storage, since it has been proved (6) that a universal computer (Turing machine) needs no more than one binary digit of internal storage (when its input-output tape is of unbounded extent, and the tape alphabet large enough). Of course such a computer would be intolerably slow, but that is beside the point.

Notes

(*) British Journal for the Philosophy of Science, 1967, 18, pp. 144-147. back

(1) I. J. Good, *Computers and Automation*, **8** (1959), 14-16 and 24-26. back

(2) J. R. Lucas, Minds, Machines, and Godel, *Philosophy*, **36** (1961), 112. back

(3) I. J. Good, *New Scientist*, **26** (1965), 182-3, and letters in *New Scientist*, 27 may and 26 August, 1965. back

(4) K. Gödel, *Monatshefte für Mathematik und Physik*, **38** (1931), 173. English translation by B. Meltzer published by Oliver and Boyd, Edinburgh and London, 1962. back

(5) I assume here an 'infinite axiom' of truth: that the conjunction of any number of true propositions is true. back

(6) E. C. Shannon, in *Automata Studies* (Princeton University Press, 1956), p. 157. back