THE SIGNIFICANCE OF INCOMPLETENESS THEOREMS *

R. L. GOODSTEIN

I WANT to start by considering certain fundamental differences between the major incompleteness theorems which have been discovered in researches in the foundations of mathematics during the past thirty years and the incomplete axiom systems which were found in the study of projective geometry during the last century. A confusion between these forms of incompleteness has led some mathematicians to underestimate the significance of the newer results and some philosophers to seek to understand the meaning of the new discoveries by reference to the technically simpler older work.

When it was discovered, for instance, that a system of projective geometry in two dimensions which postulated the axioms of incidence and Pappus' Theorem was incomplete because Desargues' Theorem on perspective triangles was not derivable in the system, then this incompleteness could be interpreted as a proof of the independence of Desargues' Theorem, postulated as a new axiom, from the other axioms. The impossibility of proving Desargues' Theorem is *surprising* in view of the fact that the corresponding theorem in three dimensions is derivable from the three-dimensional axioms of incidence, but it is nevertheless, I think, without philosophical significance because it throws no light on the nature of formal systems as such and imposes no limitations upon the axiomatic method.

The great modern incompleteness theorems which I shall consider are those due to Skolem and Gödel. Skolem's incompleteness theorem was discovered in an attempt to explain a paradox which Skolem himself found in the theory of sets.

The paradox out of which Skolem's incompleteness theorem arises, is produced by applying a result of Löwenheim's to a formalised set theory. Löwenheim showed that every consistent set of statements has a denumerable model, and so any formal system which admits some model (of the power of the continuum perhaps) has also a *denumerable* model. That is to say, for a consistent theory, we can find an interpretation in which all the objects, of which the theory

* Received 25.vii.62

treats, may be taken to be the natural numbers. Consider now some formalisation of set theory; according to the Löwenheim theorem we can find an interpretation of the membership relation of the theory in which all the *sets* of the theory are taken to be natural numbers. But in any adequate formalisation of set theory, using the familiar diagonal process, we can prove Cantor's theorem that the set of all subsets has a *greater* cardinal than the set itself:

Let S be a denumerable set and let a subset of S be denoted by a sequence of zeros and units, a zero in the nth place showing that the nth member of S is *not* in the subset, and a unit in the nth place showing that the nth member of S is a member of the subset. Suppose now that the set of all subsets of S *is* denumerable, and let it be enumerated as follows:

$$s_1 = a^1{}_1, a^1{}_2, a^1{}_3, \ldots$$
$$s_2 = a^2{}_1, a^2{}_2, a^2{}_3, \ldots$$
$$s_3 = a^3{}_1, a^3{}_2, a^3{}_3, \ldots$$

where each $a^n{}_r$ is either 0 or 1.

Define $$b_n = a^n{}_n + 1 \pmod 2$$

and consider the subset

$$\sigma = b_1, b_2, b_3, \ldots$$

The subset σ differs from S_1 in respect of the first element of S since b_1 is 1 or 0 according as $a^1{}_1$ is 0 or 1, from S_2 in respect of the second element, and so on. Thus the subset σ does *not* occur in the enumeration $S_1, S_2, S_3, \ldots$ and the hypothesis that the set of all subsets of S is enumerable is disproved.

But by the Löwenheim theorem there is a model of set theory (supposed consistent) in which each set is associated with a natural number, so that in defiance of Cantor's theorem, the set of subsets *is denumerable.*

This is Skolem's paradox. The conclusion which Skolem himself drew from the paradox is that a formalisation of set theory can contain only *relatively non-denumerable* sets; i.e. sets which are non-denumerable only because the formalisation lacks the functions to enumerate them. In other words every formalisation of set theory must be incomplete in the sense that there are denumerable sets which cannot be *proved* denumerable within the theory. To justify this interpretation of the paradox one must observe that the proof of Cantor's theorem in some system starts by assuming that a certain mapping of a set on its subsets *exists* and derives a contradiction *in the system* from this assumption. Existence here of course means existence in the system, so that the conclusion to be drawn from the contradiction is, not that the mapping in question does not exist, but only that it does not exist *in the system*. Since Löwenheim's theorem assures us that the mapping does in fact exist, it follows that *there is a mapping which is not contained in the formal system*, so that the system is incomplete with respect to the class of mappings it contains. Unlike the situation in projective geometry we cannot remedy the deficiency by fortifying the system with another axiom; we could complete the system only at the price of rendering it inconsistent. Another way of expressing the result is to say that Löwenheim's theorem for any particular formalisation of set theory is not provable by means of the resources of that formalisation alone.

Of course the notion that there are no absolutely non-denumerable sets is not a new one. The sole ground we have for believing in the existence of a non-denumerable set lies in Cantor's theorem itself. But if we do *not* assume that the totality of subsets forms a set (and this is nothing *but* an assumption) then all that the diagonal process proves is that from *any* sequence of subsets we can construct another subset, just as from any natural number we can construct another, by adding one. And if we give up the axiom of subsets of course the Skolem paradox disappears.

There is nothing in the paradox itself to *force* us to give up the axiom of subsets. A constructivist who already rejects the set of all subsets on other grounds will not need to reckon with the Skolem paradox; and a mathematician seeking the greatest possible generality will have to remain content with *relative* non-denumerability. At best we can have a transfinite hierarchy of systems in each of which there are sets non-denumerable in a particular system but denumerable in a system of greater ordinal.

The Gödel Incompleteness Theorem

I come now to the major incompleteness theorem of mathematical logic, Gödel's Theorem that all sufficiently rich formal systems necessarily contain sentences which are neither provable nor refutable in the system. There are so many interesting facets to this result that I shall later consider the proof in some detail, but first I want to observe that in this theorem, as in the Skolem theorem for sets, the incompleteness revealed by the theorem cannot be filled by means of a new axiom; it is true that the particular undecidable sentence constructed in the proof can itself be postulated as an axiom, but the proof shows that the system so fortified will still contain undecidable sentences. The undecidable sentence in this Theorem of Gödel has the form $(\forall x)\ P(x)$; neither $(\forall x)\ P(x)$ itself, nor its negation $(Ex) \neg P(x)$ is provable in the system under consideration, system $\mathscr{F}$ say, but each instance of the universal sentence $P(x)$, namely $P(0), P(1), P(2), \ldots$ *is* provable in $\mathscr{F}$. Many attempts have been made to close the gap this theorem is thought to reveal in the proof structure of formal systems. The most obvious course to take would be to add to the proof resources of the system a new derivation scheme permitting the derivation of $(\forall x)\ P(x)$ from the infinite sequence of sentences

$$P(0), P(1), P(2), \ldots;$$

but this device entirely destroys the finite character of a proof process. In the Gödel Theorem as we shall see, the very possibility of *proving* an infinite number of sentences $P(0), P(1), \ldots$ without a prior proof of the universal sentence $P(x)$, was revealed for the first time. This aspect of Gödel construction has been cleverly exploited in a recent attempt to obtain a (relatively) closed proof system without introducing non-finitist proof schemata, and is in some respects one of the most interesting features of the Gödel Theorem.

As is well known, the heart of Gödel's construction is a one-to-one mapping of the syntax of a formalised arithmetic upon arithmetic itself. There are many ways known in which this mapping may be accomplished but I shall simply suppose that each primitive sign of some formal system of arithmetic $\mathscr{A}$ has been assigned a number, and that each sequence of primitive signs with numbers $n_0, n_1, n_2, \ldots, n_k$, whether forming a sentence or not, is given the number $2^{n_0} . 3^{n_1} \ldots . p_k^{n_k}$, where p_k is the kth odd prime number. I shall further suppose that the formal system $\mathscr{A}$ contains all primitive recursive functions, either directly in the sense of admitting primitive recursive definitions as axioms or indirectly by having definition resources like existential and minimal operators. I may mention in passing that it is the failure to introduce this requirement and to show the fundamental part it plays, which vitiates most popular accounts of Gödel's work. I shall also assume that $\mathscr{A}$ is recursively axiomatisable, so that the predicate 'n is the number of an axiom' is expressed by a primitive recursive relation $A(n)$, i.e. is expressed within $\mathscr{A}$ itself by this relation. Such relations as 'n is the number of a one variable primitive recursive function', 'n is the number of a variable in formula number f', 'n is the number of a proof of formula f' and 'n is the number of the formula which results by substituting the numeral representation of number k in formula f' may all be shown to be primitive recursive. The key tools in establishing these results are the primitive recursiveness of the relation

$$(Ex)\{x < y \ \& \ R(x, y, z)\}$$

where R itself is primitive recursive, and the reductions to primitive recursive definition of a related schema of definition known as definition by course-of-values recursion. To exhibit Gödel's undecidable sentence I denote by

$$St_f(\nu/r)$$

the number of the expression of the formal system $\mathscr{A}$ obtained by substituting the numeral representing the number r for the variable of number ν in the formula with number f, and by

$$\mathrm{Pr}(m, n)$$

the relation which says that m is the number of a proof of formula number n. Further let μ be the number of the variable n and let $\mathbf{a}$ be the number of the sentence of A which we are denoting by

$$(\forall m) \neg \mathrm{Pr}(m, St_n(\nu/m)) \qquad \text{(i)}$$

and finally let $(\forall m)$ G(m) denote the sentence obtained from (i) by substituting the numeral $\mathbf{a}$ for the variable n. Then neither $(\forall m)$ G(m) nor its negation is provable in $\mathscr{A}$. We observe first that since $(\forall m)$ G(m) is formed by substituting numeral $\mathbf{a}$ for the variable n in formula number $\mathbf{a}$, its number is therefore

$$St_{\mathbf{a}}(\nu/\mathbf{a}).$$

Hence if $(\forall m)$ G(m) were provable, and if $\mathbf{k}$ were the number of its proof, then

$$\mathrm{Pr}(\mathbf{k}, St_{\mathbf{a}}(\nu/\mathbf{a})) \qquad \text{(ii)}$$

holds (and the formula in A which this represents is provable in $\mathscr{A}$, *because* Pr(m, n) is a primitive recursive relation, and this is one of the points where this fact is critical to the proof); but this contradicts the formula which $(\forall m)$ G(m) itself represents, i.e. that represented by

$$(\forall m) \neg \mathrm{Pr}(m, st_{\mathbf{a}}(\nu/\mathbf{a}).$$

I emphasise again that the contradiction is in $\mathscr{A}$ itself, and so if $\mathscr{A}$ is consistent then $(\forall m)$ G(m) is not provable in $\mathscr{A}$. To prove that $\neg$ $(\forall m)$ G(m) is also unprovable I shall assume rather more than the consistency of $\mathscr{A}$, the so-called ω-consistency of $\mathscr{A}$, but this additional assumption could be dispensed with at the price of taking a rather more complicated sentence than $(\forall m)$ G(m). By the ω-consistency of $\mathscr{A}$ we mean that for any formula $\mathscr{G}(m)$ it is impossible to prove in $\mathscr{A}$ that

$(Em) \neg \mathscr{G}(m)$ and $\mathscr{G}(0)$, $\mathscr{G}(1)$, $\mathscr{G}(2)$ . . . all hold simultaneously. First we observe that as a consequence of what we have already established, if $\mathscr{A}$ is simply consistent, then none of the numbers 0, 1, 2, . . . is the number of a proof of formula number $St_a(\nu/a)$, and so

$$\neg \; Pr(m, St_a(\nu/a))$$

is provable in $\mathscr{A}$ for any m, i.e. G(0), G(1), G(2), . . . are all provable. Hence by ω-consistency, $(Em) \neg G(m)$ is *not* provable. Thus we have seen that neither $(\forall m)$ (Gm) nor its contrary is provable in $\mathscr{A}$, although each instance of the general formula, viz. G(0), G(1), G(2), . . . is provable. The formula $(\forall m)$ $G(m)$ is said to be undecidable.

I remarked earlier that Gödel's arithmetisation showed for the first time how it is possible in a formal system with finite proof procedure to prove all the formulae G(0), G(1), . . . *without first proving* $G(m)$, (in fact even if $G(m)$ is not provable); we have just seen an instance of this. In the general case, let $N(k)$, be the number of the numeral representing k in A, and let h be the number of a formula $H(n)$, then the assertion that *all* the formulae H(0), H(1), . . . are provable is expressed by

$$(\forall k)(Em) \; Pr(m, St_h(\nu/N(k))$$

and this formula may be provable in $\mathscr{A}$ even though $H(n)$ is not. This constitutes a *formalisation of the notion of an arbitrarily assigned integer.*

The specific instance of an undecidable formula $(\forall m)$ $G(m)$ which is constructed in Gödel's proof is of no particular significance in arithmetic, but by formalising the proof of undecidability Gödel obtained the remarkable conclusion that the sentence of arithmetic which entails arithmetic's freedom from contradiction, viz.

$$(\forall m)(\forall r)(\forall s) \neg \{P(r, m) \;\&\; P(s, \text{Neg } m)\} \qquad (C)$$

(where Neg m is the number of the negation of sentence number m) is itself undecidable, if arithmetic is consistent. Contrary to a widely held belief this result does not, however, establish the impossibility of proving the consistency of a codification of arithmetic by finitist methods formalisable within the codification. Even though the closed formula C is not provable in $\mathscr{A}$, each of its instances

$$\neg \; P(\mathbf{r}, \mathbf{m}) \;\&\; P(\mathbf{s}, \text{Neg } \mathbf{m})$$

is provable in $\mathscr{A}$ for arbitrary $\mathbf{r}$, $\mathbf{s}$, $\mathbf{m}$ (in virtue of Gentzen's consistency proof by transfinite induction), and the general formula expressing the provability of these instances may itself be provable in $\mathscr{A}$. But of course a proof inside $\mathscr{A}$ of $\mathscr{A}$'s consistency offers no security, for if $\mathscr{A}$ were inconsistent then every formula in $\mathscr{A}$ would be provable in $\mathscr{A}$.

In fact, Gödel's result does not really bear upon the problem of consistency itself but affords a means of establishing the independence from the axioms of $\mathscr{A}$ of axioms (like transfinite induction) whose addition to $\mathscr{A}$ suffice to prove the closed formula C above in the enlarged system.

If we add the unprovable formula $(\forall m)\ G(m)$ to $\mathscr{A}$ as a new axiom forming a system A^+ say, then exactly as before we can construct an undecidable formula in A^+, so that A^+ is also incomplete; even if we form a new system B, by adding in all the undecidable formulae of $\mathscr{A}$ as additional axioms, if the axioms of B form a recursive set, then B is still incomplete. Since $(\forall m)\ G(m)$ is not provable in $\mathscr{A}$ the system A^- formed by adding the *denial* of $(\forall m)\ G(m)$ is consistent if $\mathscr{A}$ is consistent, and so by Löwenheim's Theorem A^- must admit a denumerable model in which G(0), G(1), . . . are all true but $(\forall m)\ G(m)$ is false and therefore m must take values other than 0, 1, 2, . . .

It follows from this that A^- must admit what is called a non-standard model, that is, an interpretation in which a class of objects which is not ordinally similar to the natural numbers plays the part of the natural numbers. That this in fact is the case was shown independently by Skolem in 1934, who proved that a certain class of functions can play the number role.

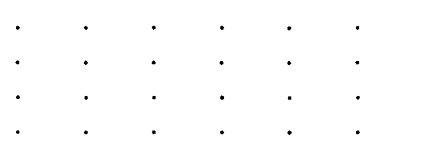
The construction of the formula $G(m)$ above can also be carried out in a system without quantifiers, in some formalisation of recursive arithmetic, $\mathfrak{R}$ say, and we find that $G(m)$ with free variable m is unprovable in $\mathfrak{R}$, but each of G(0), G(1), . . . is provable. We cannot therefore explain away Gödel's incompleteness theorem as a defect of quantification theory. $\mathfrak{R}$ like A is also incomplete. There is however an important difference between $\mathfrak{R}$ and $\mathscr{A}$ since it can be shown that no non-standard model of $\mathfrak{R}$ can itself be a recursive model.

Discussing Gödel's incompleteness theorem in 1934, before Skolem's result was known here, Wittgenstein was led to the same interpretation of the theorem, that induction and substitution of natural numbers for free variables fail to ensure that the natural numbers are the *only* values which the variables may take; it is perhaps surprising that the passages on Gödel's Theorem in the recently published 'Reflections on the Foundations of Mathematics' give no hint of this remarkable insight.

Since every axiom system for the natural numbers is incomplete and therefore necessarily admits a non-standard model it has been argued that we must look outside axiom systems for a logical foundation of arithmetic; this view is associated with a neo-realist outlook in foundation studies, that the elements of mathematics are objects in a real world with so-to-speak physical properties which the mathematician only partially captures in an axiom system. Certainly in some of his work the mathematician has an almost overwhelming conviction that he is uncovering connections which lie waiting to be revealed, but this may only reflect his amazement at the astonishing way the pieces of a puzzle sometimes fit together; we feel that the pieces must have been made to fit before we actually handled them, that we are only reconstructing a puzzle some other mind set for us. Neo-realism is not, however, a return to the Greek standpoint that formal geometry is an account of the space of our physical sensations. The case against classical philosophical realism in mathematics is overwhelmingly strong. In geometry the well-known consistency proofs of non-Euclidean geometry relative to the Euclidean makes it impossible for only one of the two geometries to be valid and yet both cannot mirror the real world. The neo-realist argues from this, not that the elements of geometry are *concepts*, but that no formal system can adequately express the whole of *geometry*, which is something revealed only to the intuition. This is an attractive thesis; every mathematician is conscious of possessing an 'inner sight', an inward short cut. But as a philosophical analysis of mathematics, neo-realism is no more tenable a position than classical realism. Intuition can be false and misleading and the inward 'short cut' just a *cul de sac*; intuition is certainly an important element in the creation of mathematics, but to see it as the organ which gives the mathematician access to mathematical reality is to be deceived by an analogy. When we say that no formal system can characterise the number concept, we do not mean that the number concept is something which we already have independently of the formal system; I may reject every definition of the meaning of a word, because it fails to characterise what I mean by the word, and maintain, rightly, that I know well what the meaning is, and yet my knowing what the meaning is may consist in nothing more than my rejection of the definitions. Just as I may write a story and be left with the feeling that this is not the story I meant to write, although of course I have not already in mind another story with which I compare it. When we contrast formal mathematics with intuitive mathematics we are not contrasting an image with reality, but a game played according to strict rules with a game with rules which change with the changing situation; a proof in intuitive mathematics may be a particular way of looking at a diagram, i.e. a particular way of using it as a symbol, or it may consist in stepping outside the particular system with which we are operating.

What we call an intuitive proof of some particular mathematical relation, is not a proof intelligible only to some special sense, quite the contrary. An intuitive proof of the relation is a proof which makes the minimum appeal to esoteric knowledge, which links the relation most immediately to a familiar background; but in its role as proof an intuitive proof has the same essential character as a formal proof, it exposes the connections between one relation and another.

An intuitive proof may for instance be a proof in which generality is expressed without the use of variables. For instance I may prove the general theorem $ab = ba$ without introducing variables, by looking at the array



first as five rows of seven dots and then as seven columns of five dots. What is *general* now is the *method* of proof; to show some one that the proof *is* general, it may be necessary to write it out again, with different numbers of dots, but this only means that we seek to draw attention to certain features of the proof, not that the proof appeals to a different *sense* than a proof which uses symbols for generality. The proof is just as formal as the proof with variables and quantifiers. What a single formal system is unable to do is to comprehend all possible partial systems in a single whole—only in this sense are formal systems necessarily incomplete; the only ' reality ' with which we can contrast a formal system is another system in a hierarchy of more or less uniformly formalised systems.

There is another incompleteness theorem for $\mathscr{A}$ which has no parallel in $\mathscr{R}$. It can be shown (for instance by means of the well-known result of classical analysis that a bounded monotonic increasing sequence is convergent) that there is a formula F(x) in $\mathscr{A}$ such that (Ex) F(x) is provable in $\mathscr{A}$, but none of F(0), F(1), F(2), . . . is provable. The formula F is this incompleteness theorem itself contains a quantifier, and no example has yet been found of a primitive recursive predicate R such that (Ex) R(x) is provable (in some formalisation of arithmetic) and yet none of R(0), R(1), R(2), . . . is provable.

Gödel's construction of an undecidable sentence in a formal system $\mathscr{A}$ utilises a detailed knowledge of the proof procedure of the system and the undecidability may be thought to reflect this structure. Kleene has however given a uniform process for finding an undecidable formula in every suitable consistent formal system. By means of Gödel arithmetisation one may determine a primitive recursive relation $T(z, x, y)$ such that for $z = 0, 1, 2, \ldots$, $(Ey)\ T(z, x, y)$ enumerates (with repetitions) all relations of the form $(Ey)\ R(x, y)$ with general recursive R. Hence given any general recursive R we may determine r so that

$$(Ey)\ R(x, y) = (Ey)\ T(r, x, y).$$

Let S be a formal system in which every general recursive function may be expressed and evaluated and such that for any formula F(x)of S there is a general recursive relation $R_F(x, y)$ which holds only when y is the number of a proof of F(x) in S.

Let $t(x, y)$ be the representation in S of the primitive recursive relation $T(x, x, y)$, let $A(x)$ stand for $(\forall y) \neg\, t(x, y)$, and let r be the number given above such that

$$(Ey)\ R_A(x, y) = (Ey)\ T(r, x, y) \qquad (3)$$

Then if S is consistent and ω-consistent, neither $\forall(r)$ nor $\neg\, A(r)$ is provable in S.

For if $A(r)$ is provable, let p be the number of its proof, so that $R_A(r, p)$ holds, whence by (3) there is an η such that $T(r, r, \eta)$ holds, and therefore $t(r, \eta)$ is provable; but if $A(r)$ is provable, $\neg\, t(r, \eta)$ is provable. This contradiction in S shows that $A(r)$ is not provable in S, and therefore $\neg\, (Ey)R_A(r, y) \equiv \forall(\eta) \neg\, T(r, r, \eta)$ holds; consequently $\neg\, t(r, y)$ is provable for each y, and therefore $(Ey)\ t(r, y)$ is unprovable, i.e. $\neg\, A(r)$ is unprovable.

Kleene's procedure can be applied only to systems with quantifiers and in this respect is less general than Gödel's original construction which is applicable also to a free variable system. It is perhaps also worth noting that even in Kleene's procedure the actual instance of an undecidable sentence is a function of the formal system being considered, since the constant r in $A(r)$ is determined by the proof predicate $R_A(x, y)$ and this of course varies from system to system. But the most interesting feature of the proof is the contrast of the semi-formal predicate $T(x, x, y)$ with its intended formal counterpart $t(x, y)$. In the theory of the predicate $T(x, x, y)$ we suppose we have before us a system of equations from which the value of a function $f(x)$ is derived by repeated substitution, a certain incompletely defined auxiliary function in the equations yielding a value only when we reach a value of y such that $R(x, y)$ holds; we then use Gödel numbering to replace the syntactical notion of deriving from a system of equations by a primitive recursive function (which we may for present purposes identify with) $T(r, x, y)$, r being the number of the system of equations and y the number of the derivation of an end equation $f(x) = \zeta$ from the system, the final step is to link $R(x, y)$ and $T(r, x, y)$ by the observation that there is a derivation of an end equation if and only if there is a y for which $R(x, y)$ holds. We are not concerned here with a formula in a formal system

$$(\mathrm{E}y)\, R(x, y) = (\mathrm{E}y)\, T(r, x, y)$$

but, as we say, with an assertion of *existence*. What does this mean? In this case that we have a procedure that enables us to *construct* the y for which $T(z, x, y)$ holds from a knowledge of the y for which $R(x, y)$ holds and conversely. But of course this *procedure* is a purely formal procedure. The situation is exactly akin to a familiar application of mathematics. If oranges cost 3d. each then I must pay 1/3d. for five; the reason for paying 1/3d. is the formal equation

$$5 \times 3 = 12 + 3,$$

but the actual purchase and payment lie outside the formal system. Depicting a computation procedure in this way (without necessarily using it) is one of the things we mean by an intuitive proof.

It is often said that Gödel's formula $(\forall m)\, G(m)$ is *true* but unprovable. The reason for saying that it is true is presumably that since each of $G(0)$, $G(1)$, $G(2)$, . . . is provable, and so true, therefore $G(m)$ is true for all m, which is just another way of saying that $(\forall m)\, G(m)$ is true. Of course if we do mean nothing more by saying that $(\forall m)\, G(m)$ is true than that $G(m)$ is true for all m then it is certainly true to say that $(\forall m)\, G(m)$ is true but unprovable. But the expression is a rather misleading one. The relationship between the formal system and the metalanguage which is established by recursion assures us that if $R(m)$ is a primitive recursive predicate such that $R(\mathbf{m})$ holds for some $\mathbf{m}$ then certainly $R(\mathbf{m})$ is provable in $\mathscr{A}$, or rather the formula in $\mathscr{A}$ which represents $R(\mathbf{m})$ is provable. But it is the essence of Gödel's theorem itself that although $G(m)$ is primitive recursive the formula $(\forall m)\, G(m)$ does *not* express the notion 'for all m, $G(m)$' in the formal system. As we have seen there is an interpretation of the system in which $G(0)$, $G(1)$, $G(2)$, . . . are *not* all the instances of $(\forall m)\, G(m)$ and therefore the truth of these instances is not to be identified with the truth of the formula $(\forall m)\, G(m)$.

Another common mistake is to suppose that $(\forall m)\ G(m)$ is true because it truly affirms of itself that it is non-demonstrable. We recall that $G(m)$ is an abbreviation for the formula which is obtained by substituting the number **a** of the formula

$$(\forall m) \neg \Pr(m, St_n(\mathbf{v}/n))$$

for the variable n in this formula.

The number of the resulting formula is of course

$$\mathrm{St}_{\mathbf{a}}(\mathbf{v}/\mathbf{a})$$

as we already have had occasion to remark; hence the fact that $(\forall m)$ $G(m)$ is unprovable, i.e. that formula number $\mathrm{St}_{\mathbf{a}}(\mathbf{v}/\mathbf{a})$ is unprovable, tells us that

$$\neg \Pr(m, \mathrm{St}_{\mathbf{a}}(\mathbf{v}/\mathbf{a}))$$

is provable for each value of m, but this of course tells us nothing about the formula

$$(\forall m) \neg \Pr(m, \mathrm{St}_{\mathbf{a}}(\mathbf{v}/\mathbf{a}))$$

that is, nothing about $(\forall m)\ G(m)$. The Gödel numbering establishes a code in which each instance of the numerical formula $G(\mathbf{m})$ says that $\mathbf{m}$ is not the number of the proof of $(\forall m)\ G(m)$, but $(\forall m)\ G(m)$ itself says nothing at all in the code. Thus the Gödel sentence is neither an example of self-reference nor of self-description. Even the sense in which we can say that the formula of $\mathscr{A}$ which we are denoting by

$$\Pr(\mathbf{x}, \mathbf{y}) \qquad \text{(P)}$$

says that $\mathbf{x}$ is the number of the proof of formula number $\mathbf{y}$ is in need of clarification. As a formula of $\mathscr{A}$, P says nothing at all. Asthe representative in $\mathscr{A}$ of a certain arithmetical relation it says that y is the exponent of the greatest power of the greatest prime number which divides x; and only as a sentence of the code which the Gödel numbering establishes, does this arithmetical relation say that $\mathbf{x}$ is the number of the proof of formula number $\mathbf{y}$.

Even supposing, which is not in fact the case, that there is a formula of the formal system (let us call it ϕ) such that *as a sentence of the code* ϕ says something about the formula ϕ of the formal system, we still could not claim that ϕ is an example of successful self-reference or self-description, for as an element of the formal system ϕ is just a sign pattern, and as a sentence of the code ϕ refers not to itself i.e. not to its meaning, but to the sign by which it is expressed, in the way the sentence

'This is written in chalk'

refers to its physical character, not to its sense.

What a sentence affirms depends upon the language in which the sentence is being used as a sentence, and identical sentences may express different propositions in different languages, as the obvious example of a code in which every English sentence stands for its contrary shows. But we must not, therefore, suppose, as many logicians do, by analogy with the specification of the range of a variable in mathematics, that every sentence *p* must be qualified by another sentence which names the language to which *p* belongs; for this assumption leads to an infinite hierarchy of languages, without achieving its aim. Whether we are speaking a common language or not cannot ultimately be settled by language alone but must show itself in our actions.

The University,
Leicester