# *12. The Philosophical Significance of Gödel's Theorem* (1963)

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: the statement being of the form $\forall x\, A(x)$ with $A(x)$ a decidable predicate. If this way of stating Gödel's theorem is legitimate, it follows that our notion of 'natural number', even as used in statements involving only one quantifier, cannot be fully expressed by means of any formal system. The difficulty is to assess precisely the epistemological significance of this result.

A common explanation is as follows. Since $U$ is neither provable nor refutable, there must be some models of the system in which it is true and others in which it is false. Since, therefore, $U$ is not true in *all* models of the system, it follows that when we say that we can recognise $U$ as true we must mean 'true in the *intended* model of the system'. We thus must have a quite definite idea of the kind of mathematical structure to which we intend to refer when we speak of the natural numbers; and it is by reference to this intuitive conception that we recognise the statement $U$ to be true. On the other hand, we can never succeed in completely characterising this intuitive conception by means of any formal system, that is, by any finitely stateable stipulation of the set of statements about natural numbers which we are prepared to assert.

On this view, then, we have a certain, quite definite, concept, which cannot be fully characterised just by the fact that we make certain assertions about it. Perhaps, indeed, it could be characterised by the fact that we make these assertions, taken together with the fact that these assertions have the meanings which they do; but in this case this latter fact could not in turn be exhaustively explained just by reference to what assertions we make. In





attempting to characterise the totality of natural numbers—e.g. for the purpose of pointing out that any model of our formal system in which $U$ is false will contain elements not in this totality—we should use some expression such as 'set' or 'finitely often'. If we then tried to give an account of the meaning of this expression by means of a stipulation of the assertions we wish to make involving *it*, this account could in turn be embodied in a formal system within which our definition of 'natural number' could be given. In view of the fact that Gödel's theorem applies to any system which contains arithmetic, there would again be an arithmetical statement expressible but not provable in this system, which we could recognise to be true: we should thus not have succeeded by this means in giving a complete characterisation of the concept 'natural number'.

Those who accept this view of the matter readily draw the conclusion that the expression 'natural number' is a counter-example to the thesis that the *meaning* of an expression is to be explained in terms of its *use*. In this connection, the application of number-words—their use to answer the questions 'How many?' and 'How often?'—is unproblematic; since we know how to give a satisfactory account of this aspect of their use, we need concern ourselves only with purely arithmetical statements. And just what we cannot do, on this view, is to characterise completely the meaning of 'natural number' by specifying which arithmetical statements we are prepared to assert and which forms of inference within arithmetic we are prepared to accept. Everyone who is familiar with the expression 'natural number' has a perfectly clear intuitive grasp of its meaning; but its meaning is such (on this view) that no account of our—or any possible—use of this expression can exhaustively explain what it is for it to have that meaning.

A natural objection is that, since I cannot look into another man's mind in order to read there what meaning he attaches to 'natural number', since all I have to go on is the use which he makes of this expression, I can never know for certain that he attaches to it the same meaning as I do. Thus a place is left for scepticism of a kind closely related to that expressed by one who asks: 'How do I know that what we both call "blue" does not look to you as what we both call "red" looks to me?' The difference between the two cases is that the supposed divergence in the private meanings of 'blue' can never come to light, whereas a divergence in the meanings attached to 'natural number' may come to light; it remains that, on this account, nothing anyone may say can ever guarantee that he means the same by 'natural number' as I. But someone who is attracted by this account may be willing to treat the possibility of such scepticism lightly; he will simply affirm that he *knows* that we all have the same concept of 'natural number'.

Evidently a large part of the interest of this question derives from its connection with the general problem of meaning, and of the relation between meaning and use. Words, as Aristotle said, are signs of ideas. We use sentences to communicate thoughts: because we cannot transmit the thoughts directly, we have to code them by means of audible or visible signs, but what we are interested in transmitting is the thought for which the sentence is a code-symbol. So philosophers have traditionally concerned themselves, not with the sentences themselves or the words which compose them, but with the analysis of the *ideas* which constitute the senses of the words.

Recent philosophy has tended to reverse this approach. Teaching a child language is not like teaching a code. One can put a code-symbol and that for which it is a symbol side by side, but one cannot isolate the concept in order to teach the child which word to associate with that concept. All that we can do is to *use* sentences containing the word, and to train the child to imitate that use. Since we judge whether the child has learned the sense of the word by whether he uses it as we do, it seems proper to *identify* possession of the concept with the ability to make a correct use of the word (or a range of associated words). The code analogy thus drops out as misleading. Even if I can recognise in myself something which I wish to call a grasp of a concept, and which is distinct from my ability to use a given word, I cannot recognise the presence of such a thing in another person: about him I know only that he uses the word in a certain way. Hence what I recognise only in myself cannot constitute the sense of the word considered as part of our common language, but this must consist rather in the use which we all make of the word. A better analogy for the relation between words and their senses is therefore that between chess-pieces and their powers. It is true that what is of interest about a chess-piece is its power, and not its material properties: but it is not a code-symbol for that power, considered as something which could in principle exist on its own; its power consists just in the fact that we have the practice of moving it about on a board, with other pieces, in accordance with certain rules.

The general thesis that the meaning of an expression is to be identified with its use is not, indeed, particularly helpful; until it is specified in what terms the use of the expression is to be described, the thesis is merely programmatic. (For instance, the limit of total triviality would be reached if it were permissible to describe someone's use of, e.g., the word 'ought' by saying, 'He uses it to mean "ought" '.) A description of the use of a particular word will remain problematic if the meanings of the words used in the description are themselves problematic. For example, a description of the use of a word will remain problematic if of some part of it it is unclear



how to answer the question, 'How is it to be recognised that that description is appropriate?'; e.g., if the description refers to the intention of the speaker, it may be asked how we are to recognise what it is that the speaker intends. Again, of some expression occurring in the description it may be asked what consequences it has to apply that expression to something; e.g., if the use of promises is described in terms of someone's acquiring a right, it may be asked what consequences it has to ascribe the possession of a certain right to someone. I am rather doubtful whether it is possible to give any *general* characterisation of the terms in which a description of use must be given so as to be unproblematic; whether anything useful is to be attained by attempting to delineate the region beyond which further questions become futile. There is, for example, no reason to think that, for any word not expressing a simple sensory quality, there must be an informative answer to the question how we recognise that the word applies; no attempt to say how we recognise something as *funny*, for instance, has yielded any very plausible results.

There is also another decision which must be taken before the programme of explaining meaning by describing use can amount to anything very definite. Besides resolving the question in what terms a description of use must be given if it is to be genuinely explanatory, we have also to determine how we are to recognise a description of the use of a word or of a type of expression to be correct and adequate: what, in other words, *constitutes* the use of a word or expression. If we are concerned with philosophy of mathematics in general, this part of the problem is extremely intractable: we have to decide not only on what principles mathematical statements are to be judged true or false, but also what is the *point* or *interest* of the procedure of deciding on their truth-value. But in the present context, we are not concerned with this complex of problems; the issue for us is only what light is thrown by Gödel's theorem on the meaning of 'natural number' in so far as understanding its meaning involves grasping the application of the predicate 'true' to arithmetical statements.

The account, described above, of the situation revealed by Gödel's theorem certainly owes part of its appeal to the fact that it is thought to afford a counter-example to the reduction of meaning to use. The line of thought here is something as follows. It is true that we cannot lay the word and the concept side by side for the child to see that one is a code-symbol for the other: it does not follow that we must identify the concept with the use of the word. No amount of training will teach a chimpanzee to talk. We may suppose, then, that the concept is latent in the child's mind, so that, while it is the training in the use of the word which, as it were, awakens the concept, and at the same time leads the child to associate that word with it, still it may be that no finite description of the use of the word can exhaust what it is to have the concept. We all of us have the concept of 'natural number'; but no finite description of our use of arithmetical statements constitutes a full account of our possession of this concept, and this is shown by the fact that we shall always be able, by appeal to our intuitive grasp of the concept, to recognise as true some statement whose truth cannot be derived from that description of the use of such statements.

As we have acknowledged, merely to identify meaning with use does not by itself accomplish very much: but to reject the identification is to abandon all hope of an explanation of meaning, to fall back on a conception of meaning as something unanalysable; the notion of meaning will then be made to play a crucial rôle in explanations, but becomes itself quite incapable of explanation. It is true that we may sometimes be forced to plead unanalysability: for all we can tell, there may be no way of analysing the concept of something's being *funny*. But this is tolerable only when the thing in question is something whose presence we can be said to recognise. We know what it is to recognise something as funny, and we agree to a sufficient extent on what we recognise as funny. A meaning, not reducible to use, which I attach to a word, on the other hand, is something which I can recognise only in myself: I cannot recognise it in you, and I cannot tell you how to recognise it in yourself. I may indeed take it for granted that, by saying certain things to you, I can induce you to attach the same meaning to a word as I do, but I can have no evidence that my hypothesis is correct; I must rely on blind faith. Such a concept of meaning has lost all its explanatory power: since everything would be the same if there were no meanings in this sense, the hypothesis of their existence is empty. Hence, although the thesis that meaning reduces to use is in itself so thin, any apparent counter-example to it must prove to be spurious. The identification of meaning with use is a small but necessary first step: progress is to be made by asking, for each case, what use consists in and how it is to be described. To reject the identification is a retrograde move, which renders further progress impossible and induces only mystification.

It thus becomes necessary to see wherein the account given above of the significance of Gödel's theorem is mistaken, and to find an alternative account. Some philosophers, in order to preserve the identification of meaning with use, have indeed denied Gödel's theorem any but a syntactic content, holding that there is no intuitive reason for calling the undecidable statement $U$ 'true'; but their reasoning has been quite implausible. The mistake in the account we are considering lies, rather, in its misapplication


of the notion of a model. This notion has, in the context of some particular suitable mathematical theory, a quite precise content: we may talk about models of axiom-systems in set-theoretical terms, or again, as in the completeness proof for classical predicate calculus, we may talk about models defined in terms of the natural numbers and of functions on natural numbers. In such cases we are talking about the models within the framework of a mathematical theory by means of which the models can be described. The account of Gödel's theorem we are considering, however, operates with the notion of a model as if it were something that could be given to us independently of any description: as a kind of intuitive conception which we can survey in its entirety in our mind's eye, even though we can find no description which determines it uniquely. This has nothing to do with the concept of a model as that concept is legitimately used in mathematics. There is no way in which we can be 'given' a model save by being given a description of that model. If we cannot be given a complete characterisation of a model for number theory, then there is not any other way in which, in the absence of such a complete description, we could nevertheless somehow gain a complete conception of its structure.

The statement $U$ is of the form $\forall x\, A(x)$, where each one of the statements $A(0)$, $A(1)$, $A(2)$, ... is true: since $A(x)$ is recursive, the notion of truth for these statements is unproblematic. Since each of the statements $A(0)$, $A(1)$, $A(2)$, ... is true in every model of the formal system, any model of the system in which $U$ is false must be a non-standard model. On the account we are considering, this fact is interpreted as follows. We have a quite definite conception of the standard model. By means of the formal system we specify a certain set of statements as ones we can recognise to be true in that model. On closer inspection, we realise that there is a statement, not in that set, which we can also recognise to be true in the model, in virtue of the fact that whenever, for some predicate $B(x)$, we can recognise all of the statements $B(0)$, $B(1)$, $B(2)$, ... as true in the standard model, then we can recognise that $\forall x\, B(x)$ is true in that model. This fact, which we know on the strength of our clear intuitive conception of the structure of the model, we can never succeed in completely expressing in a formal system, and this is why we can never completely characterise the model by means of a formal system.

We cannot formally characterise the natural numbers up to isomorphism. Of any formal characterisation, we can describe models which we recognise as non-standard. It is, however, circular to think that, since what we mean when we speak of the natural numbers cannot be fully explained by reference to the incomplete formal characterisation, it must therefore be explained instead by reference to the conception of the standard model. For this conception must be given to us by means of some description, and this description will itself make use either of the notion of 'natural number', or of some closely related notion such as 'finite'.

There is indeed something which leads us to recognise the statement $U$ as true, and which therefore goes beyond the characterisation of the natural numbers which is embodied in the formal system. But to say that we recognise $U$ as true by recognising that a model in which it was false would be non-standard makes it look as though what we were here doing was to appeal to a principle of distinction between standard and non-standard models not incorporated in the formal system, thereby obscuring the fact that this distinction cannot be explained at all except by means of the notion of 'natural number' or some allied notion like 'numeral'. In fact, the transition from saying that all of the statements $A(0)$, $A(1)$, $A(2)$, ... are true to saying that $\forall x\, A(x)$ is true is trivial. The principle of reasoning, not embodied in the system, which we employ in arriving at the truth of $\forall x\, A(x)$, is not this transition, but rather that which leads us to assert that all of the statements $A(0)$, $A(1)$, $A(2)$, ... are true. To think of the matter in terms of models obscures this otherwise evident fact, because, by speaking of all the statements $A(0)$, $A(1)$, $A(2)$, ... as being true in some model, we slur over the gap between being able, for each numeral $\bar{n}$, to recognise that $A(\bar{n})$ is true, and being able to recognise that, for every numeral $\bar{n}$, $A(\bar{n})$ is true.

The argument for the truth of $U$ proceeds under the hypothesis that the formal system in question is consistent. The system is assumed, further, to be such that, for any decidable predicate $B(x)$ and any numeral $\bar{n}$, $B(\bar{n})$ is provable if it is true, $-B(\bar{n})$ is provable if $B(\bar{n})$ is false (the notions of truth and falsity for such statements being, of course, unproblematic). The particular predicate $A(x)$ is such that, if $A(\bar{n})$ is false for some numeral $\bar{n}$, then we can construct a proof in the system of $\forall x\, A(x)$. From this it follows—on the hypothesis that the system is consistent—that each of $A(0)$, $A(1)$, $A(2)$, ... is true.

Considered as an argument to a hypothetical conclusion—that *if* the system is consistent, then $\forall x\, A(x)$ is true—this reasoning can of course be formalised in the system. Considered as an argument for the unconditional assertion of $U$, it depends more heavily on the assumption of the consistency of the system than any piece of reasoning that can be formalised in it. In order to accept a formal proof in the system as establishing the truth of its conclusion, it is unnecessary to make the grand assumption that there are no two proofs in the system with contradictory conclusions: all that is required is to accept the truth of the finitely many axioms which appear in the proof, and the correctness of the finitely many applications of the rules of inference which are used in it. But to the argument which is supposed to establish that all of the statements $A(0)$, $A(1)$, $A(2)$, ... are true, and hence that $\forall x\, A(x)$ is true, the grand assumption of the total consistency of the system is quite essential. Therefore it is in the reasoning which shows that all of $A(0)$, $A(1)$, $A(2)$, ... are true, and not in the quite evident step from there to the truth of $\forall x\, A(x)$, that we have to appeal to something which cannot be formalised in the system: namely to the argument which is intended to show that the system possesses overall consistency.



If we interpret the meaning attached to the expression 'natural number' as being an intuitive apprehension of a model for the natural numbers, it appears impossible to deny that this meaning is completely definite. Any non-standard model—for instance one in which $U$ is false—will contain elements not attainable from 0 by repeated iteration of the successor operation. Even if we can give no formal characterisation which will definitely exclude all such elements, it is evident that there is not in fact any possibility of anyone's taking any object, not described (directly or indirectly) as attainable from 0 by iteration of the successor operation, to be a natural number; we entertain no doubts as to whether Julius Caesar is a natural number or not. The conclusion drawn from this is that there is just one standard model which we all have in mind, despite our inability to characterise it completely by formal means. My argument has been, not that there is no such one standard model, still less that there is any uncertainty about which particular objects we shall recognise as natural numbers, but that the notion of 'model' here used is incoherent. Within any framework which makes it possible to speak coherently about models for a system of number theory, it will indeed be correct to say that there is just one standard model, and many non-standard ones; but since such a framework within which a model for the natural numbers can be described will itself involve either the notion of 'natural number' or some equivalent or stronger notion such as 'set', the notion of a model, when legitimately used, cannot serve to explain what it is to know the meaning of the expression 'natural number'.

Even in the case of a finite totality, the conception of that totality is not completely characterised by the way in which an object is recognised as belonging to that totality: for two people might agree in their dispositions to recognise something as belonging to the totality, and still differ on the criteria they accepted for asserting something to be true of all the members of the totality. Still more is this true if the totality is infinite. The question whether two people mean the same by a certain expression, as it arises in everyday life, is not indeed entirely definite: no doubt there are contexts in which it would be natural to allow an agreement as to the criterion for the correct application of a predicate to imply an agreement on its meaning; in this sense, then, we do have a unique and definite meaning for the expression 'natural number'. In this case, however, an understanding of the expression 'natural number' will be insufficient to determine the criterion by which something is recognised as a ground for asserting that something is true of all the natural numbers: and it is precisely the concept of such a ground 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 extension of it. If we understand the word 'meaning' differently, 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.

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. For certain particular formal systems, we may have a genuinely informative consistency proof; e.g. for the most natural type of system, we have a consistency proof of the kind first given by Gentzen, using transfinite induction up to $\varepsilon_0$. From the consistency proof, together with Gödel's reasoning, the truth of $U$ of course follows. But in the *particular* case, we learn from this only the epistemologically unsurprising fact that the particular formal system in question fails to embody everything that we intuitively recognise as true concerning the concept of 'natural number'; e.g. in the case mentioned, the validity of transfinite induction up to $\varepsilon_0$. What needs to be explained, however, is the *general* applicability of Gödel's theorem to every intuitively correct formal system; the fact that no such system can embody all that we wish to assert about the natural numbers. 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 which we know that we can give for any formal system, provided only that it is assumed about that system that it is intuitively correct. Such a general form of consistency proof cannot, of course, be expected to be genuinely informative; it can only be the trivial kind of proof by induction on the length of formal proofs with respect to the property of having a true conclusion.

In order to carry out this proof, we have first to define the property of being a true statement of the formal system. This property is defined



inductively, simultaneously with the property of being a false statement. The definition is straightforward for quantifier-free statements, and by truth-tables for the sentential operators; a statement $\forall x\, D(x)$ is said to be true if all of the statements $D(0)$, $D(1)$, $D(2)$, ... are true, and false otherwise, while $\exists x\, D(x)$ is true if at least one of these statements is true, and false otherwise. By hypothesis the axioms of the system are intuitively recognised as being true, and the rules of inference of the system as being correct in the sense of leading from true premisses to true conclusions. Hence we may establish by an inductive argument on the length of formal proofs that each proof in the system has a true conclusion, and by another inductive argument on the number of logical constants in a statement that no statement is both true and false; concluding from this that the system is consistent.

The ordinary meaning of 'natural number' involves the validity of induction with respect to any well-defined property; the property of being true is, as applied to statements of the formal system, a well-defined property, and hence the general form of consistency proof given above is intuitively correct. In the formal system, however, we can embody the validity of induction only with respect to properties expressible in the system. Once a system has been formulated, we can, by reference to it, define new properties not expressible in it, such as the property of being a true statement of the system; hence, by applying induction to such new properties, we can arrive at conclusions not provable in it.

If, then, understanding the meaning of 'natural number' is taken as including understanding the meaning of quantifiers whose variables range over the natural numbers, the meaning of 'natural number' must, as stated above, be taken to be, not completely definite, but inherently vague. It ought here to be remarked that from the fact that a concept possesses any kind of vagueness, it cannot be inferred that there is any vagueness attaching to the notion of grasping this concept; the question whether someone understands the meaning of a certain expression may be a perfectly definite one, even though the meaning of the expression in question is itself vague. The reason why the ordinary concept of 'natural number' is inherently vague is that a central feature of it, which would be involved in any characterisation of the concept, is the validity of induction with respect to any well-defined property; and the concept of a well-defined property in turn exhibits a particular variety of inherent vagueness, namely indefinite extensibility. A concept is indefinitely extensible if, for any definite characterisation of it, there is a natural extension of this characterisation, which yields a more inclusive concept; this extension will be made according to some general principle for generating such extensions, and, typically, the extended characterisation will be formulated by reference to the previous, unextended, characterisation. We are much less tempted to misinterpret a concept possessing this variety of inherent vagueness as a completely determinate concept which we can descry clearly from afar, but a complete description of which we can never attain, although we can approach indefinitely close, than in the general case. An example is the concept of 'ordinal number'. Given any precise specification of a totality of ordinal numbers, we can always form a conception of an ordinal number which is the upper bound of that totality, and hence of a more extensive totality. For the sake of constructing any precisely formulated mathematical theory, we have indeed to choose some definite method of specifying the totality of ordinal numbers which we want to use in the development of the theory; it remains an essential feature of the intuitive notion of 'ordinal number' that any such definite specification can always be extended. This situation we are not tempted to interpret as if, in thus recognising the possibility of indefinitely extending any characterisation of the ordinals so as to include new ordinals, we were approaching ever closer to a perfectly definite ('completed') totality of all possible ordinal numbers, which we can never describe but of which nevertheless we can form a clear intuitive conception. We are content, in this case, to acknowledge that part of what it is to have the intuitive concept of 'ordinal number' is just to understand the general principle according to which any precise characterisation of the ordinals can be extended.

With a concept which exhibits the different variety of inherent vagueness possessed by the concept of 'natural number', there is a much stronger temptation to such a misconception. The reason lies in the tendency to think of the meaning of a predicate as constituted wholly by the criterion for its application. (As I have said, there is no point in engaging in a verbal dispute about the use of the word 'meaning': if the meaning of a predicate is taken in this way, what must then be acknowledged is that the meaning of quantifiers whose variables range over the extension of the predicate is not fully determined by the meaning of the predicate.) Now if the meaning of 'natural number' is thought of in this way as wholly exhausted by the criterion for recognising a term as standing for a natural number, it appears impossible to regard it as inherently vague; for, as we have seen, there is really no vagueness as to the *extension* of 'natural number'. Regarded from this point of view, the induction principle appears simply as that one of the Peano axioms which stipulates that *only* those objects which can be obtained by starting with 0 and reiterating the successor operation are to qualify as being natural numbers. If we tried to interpret 'natural number' as inherently vague with respect to its extension, we should have to regard it as the reverse of indefinitely extensible: any particular definite characterisation allows the possibility of elements in the totality not attainable from 0 by reiteration of the successor operation; by successive further formal characterisations we may exclude more and more of these non-standard elements, without ever being able to exclude them all. But if we try to think of the matter in this way, we are at once struck by the fact that the situation is not the precise reverse of that which obtains in the case of the ordinals. While it is impossible to give a precise and coherent characterisation of the totality of all objects that might be called 'ordinal numbers', it is not impossible to give such a characterisation of the totality of natural numbers; all that is impossible is to characterise that totality unambiguously. When we wish to treat of the ordinals in a precise manner, we have to settle for some characterisation of them which definitely excludes some objects which could be recognised as ordinals, whereas when we treat of the natural numbers, we do not have to settle for any characterisation of them which definitely includes some non-standard elements; there is a use of 'ordinal number' in which there are no inaccessible ordinals, whereas there is no use of 'natural number' in which there are natural numbers with infinitely many predecessors. Whereas we can describe objects for which there is no definite answer, out of context, whether the word 'ordinal' should be applied to them or not, we cannot describe an object in such a way as to leave it indeterminate whether the expression 'natural number' should be applied to it. Again, we can characterise a definite totality of ordinals—say, those smaller than the first inaccessible ordinal—without making use of the general notion of 'ordinal'; but we cannot characterise any non-standard model of arithmetic without making use of the notion of 'natural number' or some similar notion. For all these reasons, once we start off by thinking only in terms of the *extension* of the concept 'natural number', we are driven to the conclusion that it is not that the concept itself is inherently vague, but that the means available to us for precise expression are intrinsically defective; the concept itself is perfectly definite, but our language prohibits us from giving complete expression to it; the concept guides us, however, in approaching ever more closely to this unattainable ideal.

If it is acknowledged that the definiteness of the extension of 'natural number' does not imply definiteness for the notion of a ground for asserting something to be true of all natural numbers—for the meaning of quantifiers whose variables range over the natural numbers—there is no longer any reason for resisting the idea that in this respect the notion of 'natural number' is an inherently vague one. Here the principle of induction is seen in the more natural way, as a principle for asserting statements about *all* natural numbers; the inherent vagueness of the concept of 'natural number' derives from the fact that it is part of this concept that the natural numbers form a totality to which induction, with respect to any well-defined property, can always be applied, and that the concept of a well-defined property is indefinitely extensible. This is so in virtue of the fact that, once we have given a precise specification of a language, and recognised every property expressible in this language as being well-defined, we can define, by reference to the expressions of this language, another property which we also recognise as well-defined, but which we cannot express in the language.

Understood in this way, the expression 'natural number' no longer appears to furnish a counter-example to the identification of meaning with use. Language contains many expressions which exhibit a variety of types of vagueness, a sub-variety of which is that which I have here called 'inherent vagueness';[1] as we have seen, there is no reason to expect any vagueness about what is the correct description of the use of a vague expression. The use of a mathematical expression could be characterised by means of a single formal system only if the sense of that expression were perfectly definite; when, as with 'natural number', the expression has an inherently vague meaning, it will be essential to the characterisation of its use to formulate the general principle according to which any precise formal characterisation can always be extended. Such a characterisation is as much in terms of *use* as any other; there is no ground for recourse to the conception of a mythical limit to the process of extension, a perfectly definite concept incapable of a complete description but apprehended by an ineffable faculty of intuition, which guides us in replacing our necessarily incomplete descriptions by successively less incomplete ones.

There is, indeed, an obvious objection to the account which I have given. The reasoning which is intended to justify the assertion of the statement $U$ proceeded via a truth-definition for statements in the language of the formal system. It might, therefore, be urged that, by defining the property of being *true* for statements of that language, we have thereby acknowledged the completely determinate character of the concept of 'natural number', at any rate as far as statements of that language are concerned. Does not a stipulation of truth-values for statements of the system constitute a precise specification of the *senses* of those statements, whatever may be the case with the principles of *proof* governing them? It is true that there will be non-standard models for the system which agree in respect of the truth-values of statements of the system with the standard model; but, as far as statements of the system are concerned, this need not be taken as showing

[1] Cf. Wittgenstein, *Remarks on the Foundations of Mathematics*, p. 55, line 8.



any indeterminacy in *their* sense. Indeed, the truth-definition assigned the value *true* to the statement $U$, in advance of our recognising $U$ as true, and independently of whether we possessed the means to do so.

The fallacy in this objection lies in overlooking the fact that the notion of 'natural number', even as characterised by the formal system, is impredicative. The totality of natural numbers is characterised as one for which induction is valid with respect to any well-defined property, where by a 'well-defined property' is understood one which is well defined relative to the totality of natural numbers. In the formal system, this characterisation is of course weakened to 'any property definable within the formal language'; but the impredicativity remains, since the definitions of the properties may contain quantifiers whose variables range over the totality characterised. If someone chose not to believe that there was any such totality for which induction with respect to any property well defined relative to it was always valid, i.e. that we could consistently speak of such a totality, there would, as far as I can see, be no way of persuading him otherwise; obviously no formal consistency proof would be of any avail. The truth-definition, however, neither presupposes nor effects a completely determinate specification of the totality. It explains quantification over the natural numbers by means of quantification over the numerals; and these, of course, form a totality isomorphic to the natural numbers. This is unobjectionable when the truth-definition is regarded merely as defining a property which can count as well defined relative to the totality of natural numbers, and to which induction can therefore be applied: but if we try to treat it as an *explanation* of the meaning of quantifiers whose variables range over the natural numbers, it becomes useless because circular; if the meaning of quantification over the natural numbers remains to be explained, then the meaning of quantification over the numerals does too. The only way to *explain* the meanings of quantification over the natural numbers is to state the principles for recognising as true a statement which involves it; Gödel's discovery amounted to the demonstration that the class of these principles cannot be specified exactly once for all, but must be acknowledged to be an indefinitely extensible class.

I will in conclusion remark briefly on the relation of what I have said to the intuitionist philosophy of mathematics. Evidently some of the things I have said, if correct, bear out certain of the intuitionist claims. Intuitionists hold that the classical explanations of the logical constants, and of the quantifiers in particular, in terms of truth-conditions for statements in which they occur, are faulty because circular; and that the right method for explaining them, which avoids vicious circularity, is not to lay down, for each constant, conditions under which a statement for which that constant is the main operator is true, but rather to lay down criteria for recognising something to be a proof of such a statement. As far as quantifiers whose variables range over an infinite totality are concerned, this coincides exactly with what I have here asserted. This is, however, not enough in itself to compel acceptance of the intuitionist position. For the question now arises: In what terms is it permissible to state the criterion for what is to constitute a proof of some form of statement? It is essential for the intuitionists to hold that this criterion must be stated in terms of the criteria (presupposed as known) for recognising proofs of the constituent statements; just as the classical stipulation of the conditions under which a form of statement has one or other truth-value is in terms of the conditions (presupposed as known) under which its constituents have these truth-values. Nothing that has been argued in this paper goes to show that the criterion of proof has to be stated in this particular form. An intuitionist may say, indeed, that any other method of stating such a criterion must tacitly appeal once more to the illegitimate notion of determinate truth-conditions considered as obtaining independently of our methods of recognising the truth-values of statements: to consider whether there is justice in this claim would, however, involve us in considerations of a quite separate kind.

What we have been considering does bear on another intuitionist thesis: that a mathematical proof or construction is essentially a mental entity, something that may be capable of being *represented* by an arrangement of symbols on paper, but cannot be *identified* with it. This thesis is not intended merely as a protest against a superficial formalism which takes no account at all of the interpretation we put on our symbols: it is a rejection of the idea that there can even be an isomorphism between the totality of possible proofs of statements within some mathematical theory and any determinately specified totality of symbolic structures, i.e. proofs within any formal system. Intuitionist language on this matter is, rightly, repugnant to anyone who has grasped the point of Frege's repudiation of 'psychologism', of the introduction of strictly psychological concepts into logic or mathematics. But when the intuitionist conception is stripped of its psychologistic guise, it can be recognised to be entirely correct. The intuitive conception of a valid mathematical proof, even for statements within some circumscribed theory, cannot in general be identified with the concept of a proof within some one formal system; for it may be the case that no formal system can ever succeed in embodying all the principles of proof that we should intuitively accept; and this is precisely what is shown to be the case in regard to number theory by Gödel's theorem. In this case, as we have seen, this simply means that the class of intuitively acceptable proofs is an indefinitely extensible one;



but it is clear that the intuitionists are right in claiming that, if the sense of mathematical statements is to be given in terms of the notion of a mathematical proof, it should be in terms of the inherently vague notion of an intuitively acceptable proof, and not in terms of a proof within any formal system.